# GENETIC ADMIXTURE OF AQUACULTURE STOCKS INTO WILD REMNANT POPULATIONS OF THE EASTERN OYSTER, *CRASSOSTREA VIRGINICA*

A Thesis

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Master of Science

by

Yuka Kutsumi

12/2017

## ABSTRACT

Oyster reefs are one of eleven unique habitats prioritized for restoration in the Hudson River Estuary to improve overall environmental quality and provide society with new and increased benefits from the estuary environment (Miller 2013). This goal is shared by many coastal states and has inspired many supportive breeding programs to revitalize native oyster reefs (*Crassostrea virginica*) in estuaries across the Eastern United States. In some cases, these programs rely on domesticated aquaculture oyster broodstock to supplement populations with hatchery-produced oysters. However, little is known about the long-term efficacy of supportive breeding using domesticated-strain broodstock vs. adults collected from nature. One basic question is about the degree of genetic difference between these two types of oyster compared with spatial differences among natural stocks. In this study, we collected samples from different life stages (juveniles and adults) of eastern oysters from 10 different sites (N=24 per site) in the Hudson River Estuary and compared them with an aquaculture oyster strain recently used for supportive breeding. Population differentiation was tested at a genomic scale with double digest RADseq data. Among natural population samples there was subtle population structure between the Hudson and East Rivers despite the high dispersal potential present during the 2-3 week pelagic larval stage. This contrast between rivers was seen in both spat and adults, and was consistent across two sampling years. The pattern was most strongly manifest in the highest among-population Fst loci, and became recognizable as a spatially discreet pattern of introgression when the aquaculture strain was included in the analysis. The aquaculture strain showed lower genetic variation relative to wild populations. Until fitness consequences of introgression, and relative fitness of wild and aquaculture oysters can be evaluated over the full oyster life cycle, I recommend stricter adherence to published recommendations that wild broodstock be used for generating oyster restoration seed and to avoid mixed plantings that could facilitate interbreeding.

## BIOGRAPHICAL SKETCH

Yuka Kutsumi is currently in her 3rd year of study in the Department of Natural Resources at Cornell University. In August 2017, she will graduate with a Master of Science degree with a focus in conservation genetics/genomics. Yuka is a member of Badminton Club at Cornell University and will continue playing badminton for the rest of her life. Yuka received a Bachelor of Art degree in Biology from Kalamazoo College, Michigan in 2011, and she obtained a Master of Science degree in Medical and Bioinformatics from Grand Valley State University in 2013.

ACKNOWLEDGMENTS

I would like to thank my thesis advisors; Dr. Matthew Hare and Dr. Patrick Sullivan for their constant guidance and encouragement. Without their constant guidance, this work would not have been possible.

I would also like to thank the Graduate School at Cornell University and especially Randy Patterson, the health care provider at Gannet, for giving me the support I needed in a timely fashion.

Finally, I would like to thank my family, friends and office mates at Cornell University for their encouragement and moral support.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SUPPLEMENTAL FIGURES

SUPPLEMENTAL TABLES

# INTRODUCTION

## The History of Anthropogenic Impacts on the Hudson River Estuary

Human activities, directly or indirectly, are now the primary cause of changes to biodiversity ( Palumbi and Mu, 2001; Kareiva, 2014; Corlett, 2015; Lennon, 2017). The primary direct impacts are often caused from overexploitation and habitat loss, while indirect effects can result from cascading interactions in the food web (Goulletquer *et al* , 2014). Natural perturbations such as hurricanes and storms have always occurred in the aquatic systems, but the resulting changes in biodiversity were frequently seen to be reversible or have been long integrated into the larger spatial and temporal patterns of ecosystem structure and function. On the other hand, the effects of many human activities (anthropogenic impacts) are frequently irreversible (Murcia *et al* , 2014), at least over the span of a human life.

The Hudson River Estuary, located off Manhattan Island, was negatively impacted by accelerated anthropogenic activity in the 19$^{th}$ and 20$^{th}$ century. Until 1986, when the North River sewage plant started operating, 150 million gallons of raw sewage entered the Hudson River daily (Miller 2013), which led to high bacterial counts, low oxygen levels, and excess nutrients, such as nitrogen (N) and phosphorus (P), with additional contributions from agricultural runoff (Howarth and Marino, 2006). As a result, the Hudson River Estuary suffered from a range of ecological problems including increased sedimentation and turbidity, hypoxia, loss of sea grasses and suspension feeders, and a general loss of oyster reef habitat (Jackson *et al* , 2001). The estuary experienced a shift from being an ecosystem dominated by benthic primary production to one dominated by planktonic primary production (Jackson *et al* , 2001).

## Eastern Oyster (*Crassostrea virginica*) and its Population Connectivity

The eastern oyster, *Crassostrea virginica*, is native to eastern North America. This keystone species historically provided important economic and ecological services (Kennedy1996; Grizzle and Brumbaugh, 2013). Economically, eastern oyster aquaculture production in 2015 was recorded to result in 13,600 tons (meat weight) worth, $197,000,000 (USD) annual market value, of which $6,000,000 (3%) originated from NY State alone (National Marine Fisheries Service 2016). Ecologically, the ecosystem services that oyster populations provide include providing habitat for other species and improving water quality by removing nitrogen through filter feeding (Grizzle & Brumbaugh, 2013). In the Hudson River Estuary, oysters were an ecologically

important and prolific part of the ecosystem. Eastern oysters could be found from the southern limits of the Estuary through Ossining, New York to the north (Waldman 1999). At the end of the nineteenth century, oyster beds still occupied about 350 square miles of the Estuary (Waldman 1999), however, due to urban and agricultural runoff and wastewater, the oyster populations dropped to barely detectable numbers (Yozzo *et al* 2004). Oyster harvesting near New York City has now been prohibited for nearly a century and there are currently no known self-sustaining reproductive oyster reefs in around Manhattan or in the East River (Kennish, 1992; McCay, 1998; Jackson 2001), although sparse recruitment of juveniles occasionally occurs on hard substrates (M. Hare, unpublished data).

The mechanisms that define the scale and the pattern of population connectivity and the effective population size remain poorly known for eastern oysters. In marine populations with high fecundity, high early mortality and broad larval dispersal, even the dynamics of relatively pristine populations can be difficult to predict. The eastern oyster experiences a two to three-week pelagic larval phase during which dispersal between subpopulations may occur (Thompson *et al* 1996). Although the recent biophysical models have depicted the oyster larval dispersal in great detail (North *et al* 2008; Narvaez *et al* 2012), the difficulty in tracking larval dispersal under natural conditions leaves much to be understood about the fate of larvae during this period. Previous analyses of genetic markers at meso- and macro-geographic scales seem to support the view that the eastern oysters have the potential for long distance dispersal among its populations (Reeb and Avise 1990; Karl and Avise 1992; Rose *et al* 2006).

A pattern commonly seen in marine organisms, particularly those with demersal adults and pelagic larvae is known as “chaotic genetic patchiness”, wherein significant pairwise genetic differentiation occurs at spatial scales below the average dispersal distance, but the variation among sites lacks clear geographic trends and shows temporal instability (Johnson and Black 1984; Hedgecock 1994; Worm *et al* 2006). This pattern is hypothesized to result from “sweepstakes recruitment”, the regular occurrence of extremely variable reproductive success (Hedgecock 1986; Hedrick 2005; Hedgecock and Pudovkin 2011). When the variance in reproductive success is large enough, it can limit effective population size to a small fraction of the actual census population size (Turner *et al* 2002; He *et al* 2012). Empirical observations documenting this phenomenon demonstrate how extreme sweepstakes reproduction can generate cohort by cohort population structure, with each cohort seemingly produced by relatively few

contributing parents (Taris *et al* 2007; Christie 2010; Hedgecock and Pudovkin 2011). Because many proximate factors can contribute to large variations in individual reproductive success, existing theory has not been adequate for predicting when and where sweepstakes reproduction will be extreme. A thorough test for sweepstakes reproduction in Delaware Bay using microsatellite markers found a dramatic reduction of effective population size relative to census size, but no evidence of genetic patchiness (He *et al* 2012).

**Status of Aquaculture Industry in Long Island Sound**

Towards the middle of the 19$^{th}$ century, the oyster aquaculture industry was regarded as a solution to save what was left of natural stocks by reducing direct harvest pressure in the Hudson River Estuary and fill the ever-increasing market demand for oysters (MacKenzie 1996). Long Island Sound, a tidal estuary connected to New York City by the East River, hosted an expanding oyster bed culturing industry in the 1960s. This involved spreading large quantities of shell over the grounds in order to provide habitat for wild oyster larvae to settle, then transplanting these seed oysters to promote growth on the harvest grounds where mortality from starfish and oyster drill predation was expected to be lower (MacKenzie 1989; MacKenzie 1996). Oyster landings reached nearly 1 million bushels/year in the early 1990s in Long Island Sound (MacKenzie 1996). As the industry developed, producers increasingly retained and selectively bred seed oysters that they found had favorable production characteristics such as fast growth and disease resistance (Ragone *et al* 2003; Nell and Perkins 2006; Frank-Lawale *et al* 2014 ).

**Recent Oyster Restoration in the Hudson River Estuary**

Since the passage of the 1972 Clean Water Act, the water quality in New York Harbor has slowly and steadily improved (Pomeroy *et al* 2006). The 2015 assessment of the status of Hudson River Estuary by the New York City Department of Environmental Protection (Stanne 2015), reported that the bacteria levels (fecal coliform and Entercoccus counts) in the Hudson met current water swimming quality standards (60 colonies/mL). The average level of dissolved oxygen has increased, now >6 mg/L and used to be around 4mg/L, and other signs of improvement include measures of phytoplankton abundance and water clarity (Stanne 2015).

Oyster reefs are one of eleven unique habitats prioritized for restoration in the Hudson River Estuary to improve overall environmental quality and provide society with new and increased benefits in the estuary environment (Miller 2013). Apart from the oyster's culinary credentials and the desirability of a fishery, this species is viewed as a priority for restoration because of its role as an ecosystem engineer that builds reef habitat, which is used by many other species, stabilizes shorelines and helps filter particulates from the water during feeding (Coen *et al* 2007; Jackson *et al* 2001)

Motivated by the potential to restore ecosystem services and facilitated by water quality improvements, large-scale oyster restoration projects have captured the publics' imagination (Billion Oyster Project, 2013). Large scale aspirations have the potential to be realized through substantial funding from government sources (NYC DCP 2013). Pilot-scale efforts related to these restoration projects have relied on hatchery-produced oysters, mostly from domesticated-strain broodstock. Unfortunately, the long-term success rate of restoration reefs and the key determinants of restoration efficacy are not well studied (Mann and Powell 2007).

A pilot restoration project called the Oyster Restoration Research Project (ORRP), funded by the EPA and Hudson River Foundation, which monitored oyster performance and ecosystem services at 50 $m^2$ patches of replenished bivalve shell habitat at three sites in 2011-12. The unknown domesticated-strain aquaculture "seed" oysters (juvenile "spat" on shell), were planted at these sites and survival, growth, natural recruitment, and environmental conditions were monitored for two years (P. Malinowski pers. comm.). The ORRP found that the reefs at Soundview and Hastings (East River and Hudson River, respectively) showed the most favorable growth rate and natural juvenile recruitment. It was concluded that the Soundview site in the East River had the best overall environmental conditions and the greatest potential for successful oyster reef restoration (Grizzle *et al* 2013).

One risk of doing population supplementation (i.e., restoration in the context of HRE) using domesticated aquaculture strain oysters is that they might have lower lifetime fitness than wild oysters under natural conditions. Even though aquaculture oysters might grow rapidly to market size, life history theory suggests that domestication may involve tradeoffs with traits expressed at other times during the life cycle, such as fecundity, longevity or larval tolerances. If this is true then domestication can promote traits that are maladapted in the wild, possibly limiting restoration efficacy. Furthermore, interbreeding of nearby wild oysters with hatchery-produced

cohorts, if produced from domesticated strains, can potentially compromise the fitness of progeny and drain more productive wild reproductive potential (Araki *et al* 2007; Christie *et al* 2012; Eierman and Hare 2014). In theory, with sustained supplementation using cultured individuals with low genetic diversity can reduce the effective population size of the receiving wild population (Ryman *et al* 1995; Christie *et al* 2012). The consequences to the gene pool and the fitness of the population through introgression from the aquaculture strain to the wild has not been studied in eastern oysters. In this paper, we use genomic methods to investigate the amount and distribution of genetic variation among oyster populations in the Hudson and East Rivers, relative to a domesticated aquaculture strain, and test for interbreeding.

**High-throughput sequencing provides finer resolution for understanding genetic structure**

With the advancement and increasingly widespread use of next-generation high-throughput sequencing technologies, the number and type of loci available for studying marine species in both model and non-model species has increased (Van-Wyngaarden 2016). The ability to survey genome-wide diversity, including the target loci potentially associated with adaptive variation can be particularly informative in large marine populations where directional selection can drive rapid divergence and differentiation (Brandury *et al* 2010; Van-Wyngaarden *et al* 2016). One way to test for the signature of natural selection is to look for loci that show a higher degree of population differentiation than expected by drift, i.e., outlier loci. Additional patterns such as deviations from Hardy-Weinberg equilibrium can strengthen an inference of selection, but ultimately these observational methods identify candidates that must be further studied to confirm the inference (Lowry *et al* 2016). Empirical studies in several marine species have reported fine-scale geographic structure of individual loci using the outlier locus approach (Catchen *et al* 2013; Benestan *et al* 2015). Previously, Hare and Eierman (2014) assembled and annotated the *C. virginica* transcriptome, provided an opportunity in this study to potentially associate outlier loci with hypothesized function.

## OBJECTIVE AND HYPOTHESIS

In this study, our goal was to characterize the genetic variation within and among oyster populations in the Hudson River Estuary by using a reduced representation method of sampling random loci across the oyster genome. We sampled both newly-settled juveniles (spat) and adults in each of two years to test for both spatial and temporal genetic heterogeneities in the Hudson and East Rivers of New York State. In addition, because of recent plantings of hatchery-produced cohorts from domesticated aquaculture strains, we included two samples of a commonly used strain. Given the life history and biology of eastern oysters reviewed above, and the regional genetic homogeneity found previously in other portions of the species range using microsatellites and RFLPs (Reeb and Avise 1990; Karl and Avise 1992; Gaffney *et al* 1996; Rose *et al* 2006), genetic homogeneity is a reasonable equilibrium expectation at this within-estuary scale. Our null hypothesis was for genetic homogeneity among the 10 different sites or across years if the entire estuary had mixed larval dispersal and sweepstakes reproduction was weakly expressed as reported by Rose *et al* (2006) and He *et al* (2012) for *C. virginica* in other estuaries. Alternative hypotheses included chaotic genetic patchiness (Johnson and Black 1984) expressed spatially or temporally as a result of sweepstakes reproductive success. Another hypothesis was for genetic differentiation at functional loci responding to environmental gradients associated with salinity in the Hudson River. A third hypothesis was that progeny of aquaculture strains used for restoration plantings as early as 2010 (Grizzle *et al* 2013) would create some mixed genetic structures in the vicinity of those plantings.

## MATERIAL AND METHODS

### *Sample Collection*

Wild recruiting *C. virginica* spat (apx. one month post-settlement) were sampled from 10 sites in the Hudson Raritan Estuary, with additional adult samples at two sites where they could be found (Table 1, Figure1). The sites include five localities in the Hudson River and five in the East River. Four Hudson River localities were north of Manhattan where average salinity is low (0 -12 PSU) and one was near the mouth of the estuary at average salinity 22.6 PSU. East River localities had moderate salinities (20 - 26.6 average PSU). Nylon mesh bags containing clean bivalve shells were deployed in May through October of 2012 and 2013 for collecting spat recruits (two bags per site). Shell bags were checked monthly starting in July and ending in October, each time swapping in clean shell if there was fouling. Oyster adductor muscle (adults) or whole bodies (spat) were preserved in 95% ethanol.

### *Construction of the ddRAD Library*

DNA was extracted from 20 mg of adductor muscle (adults), gill tissue (large spat) or whole bodies minus gastrointestinal tract (small spat) using the DNeasy Tissue Kit (Qiagen) with RNase treatment (Thermo Fisher Scientific) for some but not all samples. DNA quality and quantity were assessed using a Nanodrop spectrophotometer and agarose gel electrophoresis. Library construction generally followed methods in Peterson *et al* (2012) and White (2013). For every individual, 30 ug of genomic DNA was normalized to a concentration of 100ng/uL and double digested using two enzymes, PstI and NlaIII. Adaptors ligated onto fragment ends included barcodes (Elshire *et al* 2011) of length 5, 6 and 7 with at least 2 pairwise nucleotide differences. Samples were distributed across seventeen multiplexed libraries (combined with additional samples for another study) and sequenced on an Illumina HiSeq 2500 platform generating 700 M single-end, 100 bp reads for the samples in this study.

### *Initial QC Filtering and Data Processing*

Adaptors (but not barcodes) were removed from reads before using custom scripts to trim from the 3' end using a sliding window of 15 nucleotides (nt) until the median base Phred score was above 33 for the entire read length. Libraries were run through STACKS *process_radtags* (ver. 1.37) using custom scripts to process each barcode length separately. Default parameters

were used except a minimum Phred33 score of 20 was required in a sliding window of 15 nt. Barcodes were then removed and each sample was scanned for both ddRAD restriction sites and their reverse compliments at trimmed read ends and internally. Internal restriction sites indicate a partial digest product and triggered discarding of the read whereas terminal partial sites were trimmed. Final 3' end trimming was used to create a uniform read length of 89bp. Additionally, we removed individuals with fewer than 0.5 million reads after the filtering process (Supplemental Figure 1).

***Pseudo-Reference Assembly, Initial Filtering and Locus Definition***

In the absence of a reference genome for *C. virginica*, we used methods in Ilut *et al* (2014) to construct a pseudo-reference with ddRAD loci defined by clustering. To avoid artifacts the optimum clustering threshold minimizes false homozygosity, caused by oversplitting alleles into different loci, and avoids the false heterozygosity created when paralogs are mistakenly clustered as alleles. Individually, using several high-coverage individuals, we collapsed quality-filtered reads down to distinct sequences having read depth ≥2. Clustering of reads was attempted iteratively for each individual with allowable sequence differences of 1 to 12, requiring a minimum of 10 reads per cluster. As allowable mismatches increased, the proportion of homozygous clusters decreased to an asymptote at 9 mismatches (Supplemental Figure 2). Then, to generate a representative pseudo-reference, we combined all sequence data (wild and aquaculture) and randomly subsampled 5 million reads, the average number of reads per sample, and clustered reads using max edit distance = 9. Each inferred cluster was represented in the pseudo-reference by the most deeply sequenced variant (i.e. most common allele).

We used Bowtie2 (ver 2.2.2) to align quality-filtered fastq reads from each individual to the pseudo-reference (Langmead and Salzberg, 2012) using the end-to-end option, which aligns reads from one end to the other without any trimming of characters from either end. All secondary alignments were removed from the resulting SAM files using the 'XS' flag. This was done because we found that end-to-end mapping combined with XS filtering generated the least number of sequences with soft-masked ends that can weaken downstream bioinformatics steps in STACKS.

***Genotyping and Quality Control***

The ddRAD loci identified for each individual by unique mapping to the pseudoreference were processed in STACKS ver. 1.37 by applying the *ref_map.pl* pipeline (Catchen *et al* 2011, 2013) to SAM files. For this pipeline, non-default parameter values included minimum number of identical reads to form a "stack"( m = 2); the upper bound on sequence error rate, bound_high = 0.01 in the genotyping model, which allowed for no mismatches (due to prior construction of a pseudo-reference). To reevaluate model calls that were not significant in each individual, Stacks *rxstacks* was applied with filtering out of catalog loci with log likelihood less than 500 (--lnl_lim -500) (Table 2). We also required that a minimum of 5% of loci in the population must be confounded relative to the catalog locus (--conf_lim 0.5). Singleton or low frequency reads were filtered using custom perl script to ensure the diploid individuals to have two alleles. The ddRAD genotypes from individuals were then combined into a catalog using cstacks and sstacks.

***Iterative STACKS Filtering***

We further filtered the STACKS catalog by iteratively using the populations module, first with all population samples combined, then to produce data sets for each of the population samples. At first, using the *populations* module (ver 1.41), polymorphic loci were only included in a dataset if they had a stack depth greater than 6 (-m6), and if they occurred in at least 98% of the individuals in a population (-p1r98) and had a minor allele frequency greater than 0.01% (–min_maf0.01) (Table2). Loci strongly deviating from the Hardy Weinberg equilibrium expectations were also removed using a whitelist based on Fis (-0.5<Fis<0.5) (Supplemental Figure 3).

***Identifying Outlier Loci with LOSITAN***

To test for loci that show more population differentiation than expected under neutrality, LOSITAN (Antao *et al* 2008) was applied to the 17 wild population samples. This method uses the observed allele frequencies to estimate the expected heterozygosity and global unbiased $F_{ST}$ values (Cockerham and Weir 1993) in order to simulate an expected neutral distribution for $F_{ST}$, assuming an island model of migration (Wright 1931). Because most of the ddRAD loci had two or more single nucleotide polymorphisms (SNP; see Results), haplotypes rather than nucleotides were used as the unit of analysis. Sixty million simulations were performed assuming an infinite

alleles mutation model. Ninety-five percent confidence intervals were drawn around the simulated mean neutral $F_{ST}$. Haplotype loci with $F_{ST}$ values significantly greater than expected under neutrality were considered as candidates for positive selection in one or more populations.

***Population Structure Analysis***

We examined population structure using a model-based method, STRUCTURE 2.3.4 (Pritchard *et al* 2000), which infers population clusters based on an assumption that populations are at Hardy-Weinberg equilibrium and in linkage equilibrium. With 17 wild population samples the model runs had a burn-in period of 100,000 iterations followed by 500,000 Markov Chain Monte Carlo (MCMC) steps for 1-6 assumed clusters (K), with three replicate runs for each K. The population admixture model was used with the location prior (LOCPRIOR) correlated, to improve the detection of weak structuring in open populations (Hubisz *et al* 2009). A separate STRUCTURE run was implemented without the LOCPRIOR setting in order to test the effect this prior had on the optimization of K. Identical parameters were used for analysis of the 17 wild plus 2 aquaculture samples. STRUCTURE outputs were post-processed using STRUCTURE HARVESTER (Earl and von Holdt, 2012), which estimated the optimal number of K using Evanno's delta-K method (Evanno *et al* 2005). STRUCTURE analyses with the 17 wild populations. The K-means clustering algorithm (Jombart *et al* 201) for $K = 1$ to $K = 10$ on the haplotype wild oyster dataset was also tested using R package (adegenet). GenAIEx 6.5 was used to conduct an analysis of molecular variance (AMOVA) with 999 permutations. The basic genetic diversity metrics were also obtained by using GenAlEx 6.5, and the unbiased heterozygosity was calculated by using the random subsample of 13 from each of the larger samples.smallest population size (N=13).

***Tests for Enrichment of Coding Regions***

To identify coding sequences and gene annotations for LOSITAN outlier loci we used NCBI BLAST to compare outlier haplotypes against the *C. virginica* transcriptome (Eierman and Hare 2016) with evalue cutoff ($1e^{-06}$). The proportion of outliers aligning to the transcriptome was compared to its expectation (the null distribution) based on blasting all 17-wild haplotype loci. We then used a randomization approach to test whether coding sequence is more common than expected among outlier ddRAD loci. We used R (R Core Team 2016) to estimate an

empirical null distribution for the proportion of ddRAD loci found to include coding sequence by randomly drawing 100,000 samples from the total number of ddRAD loci (script from Soria-Carrasco *et al* 2014). We computed the empirical cumulative distribution and calculated the two-tail p-value for the observed proportion of outliers aligning to the transcriptome.

## RESULTS

### *Locus Identification and Filtering*

Based on aligning quality-filtered reads from 349 wild individuals (17 population) to the pseudo-reference genome, 60,206 catalog loci were identified by STACKS, including 743,150 SNPs (average 9.3 SNPs per locus). After STACKS *rxstacks* filtering, we retained 9250 ddRAD loci (15.3% of the initial catalog). Further filtering by the *populations* (ver 1.41) module (loci present in more than 98% of individuals, read depth>6, minor allele frequency>0.01 over all 349 samples) produced 4212 ddRAD loci (7.0% of initial catalog). Additional Fis filtering reduced the locus count to 4186 (6.9% of initial catalog) (Table II). These ddRAD loci contained a total of 58,846 SNPs with a range of 1 to 12 per locus (average = 2). Because most loci had 2 or more SNPs, we analyzed haplotypes rather than SNP data except where noted.

When two samples of aquaculture oysters were included with the 17 wild samples. A new dataset was produced with STACKS for this 17+2 set of population samples with the same filtering applied. This dataset had 4241 ddRAD loci with a total of 22091SNPs (1 to 26 per locus).

### *Genetic Diversity*

For the final wild oyster haplotype dataset containing 349 individuals and 4186 loci, the number of alleles at each locus averaged 16.3 with a range of 2 to 108 alleles. The population sample mean number of alleles ($N_a$) was 3.155 ± 0.073 for 95% confidence intervals (CI). The mean gene diversity ($H_e$) among 17 populations was 0.366 ±0.01 (95% CI) and the mean proportion of private alleles per population sample was 0.37±0.02 (Table III). Significantly lower genetic diversity was found in aquaculture samples (FIS2012, FIS2013a) with mean number of alleles 2.124±0.046 (95%CI), the mean number of private alleles 0.080±0.01(95% CI), and the gene diversity 0.271±0.008(95%CI).

### *Population Structure and Genetic Differentiation*

We initially tested for population structure among the 17 wild population samples using the admixture model with locality prior applied to haplotype data in STRUCTURE. Evaluating K= 1 to 5 with the 17 wild population samples in STRUCTURE, K=2 was the number of populations

best supported based on the log(L) criterion (Supplemental Figure 4). K=2 was also supported by the results of the K-means clustering as well as PCA eigenvalues (Supplemental Figure 5). Instead of K=2 reflecting spatial subdivision, almost all individuals showed a similar level of mixed ancestry from the two source populations (Figure 2a). The distribution of Fst values among loci showed a very low mean (0.006) with Fst > 0.2 at relatively few loci (Figure 3). When we removed ddRAD loci with the top 10% of haplotype Fst values and reran STRUCTURE for K=1-5, the K=2 still had highest support (Supplemental Figure 3), indicating that the genetic signal generating K=2 includes many small effect loci, not just a few high effect loci. Nonetheless, we used LOSITAN to identify non-neutral loci using haplotype data and found 51 loci (1.2%) with Fst ranging from 0.032 to 0.566. Analysis of the 51 outlier loci with STRUCTURE for K=2 showed a different pattern of admixure between the Hudson River and East River samples. Hudson River samples showed genotypic variation predominantly from one source population, whereas the samples collected from East River showed roughly equal proportion of ancestry from the two source populations (Figure 2b).

To further understand the source of genotypic differentiation between the two rivers, two samples of an aquaculture strain were included in the analysis (Figure 2c). This aquaculture strain (Fishers Island) is planted annually on bottom leases by some oyster farmers in Long Island Sound and was used for pilot-scale restoration plantings at the Hastings on the Hudson and Soundview sites in the Hudson and East Rivers, respectively. STACKS population filtering was redone with 17 wild + 2 aquaculture samples to produce a dataset with 3655 haplotype loci in 392 individuals. STRUCTURE analysis of the 3655 haplotype loci for K= 1 to 6 showed similar support for K=2 or 3 models (Supplemental Figure 4d, Figure 2 c,d). In both K=2 and K=3 models, the aquaculture strain is inferred to be the source of admixture that is differentially expressed in the East and Hudson Rivers. As seen before with outlier loci from the 17 wild sample set and no aquaculture reference sample, the pattern of admixture was uniformly higher across the East River samples relative to the Hudson River samples (Figure 2c). However, with the benefit of aquaculture reference samples the average level of introgression in the East River was only 6.3% +/- 4.7% (95% CI), compared with almost no introgression inferred for Hudson River samples 2% +/- 1.8% (95% CI). Because somewhat different sets of loci were analyzed in the 17 wild versus 17+2 sample sets, the 51 outlier loci identified from 17 wild population samples were obtained from the 17+2 sample set (whitelist feature in STACKS) and analyzed

using STRUCTURE for K= 1 to 5. If the putatively non-neutral loci identified among wild population samples were a result of introgression from aquaculture oysters, then we would expect to find this admixture pattern also reflected in the 17+2 sample set and attributed to the aquaculture source. The 51 outlier loci supported K=2 (Supplemental Figure 4c) and showed a pattern of introgression from the aquaculture strain into the East River population, but not the Hudson River (Figure 2c). Note that temporal replicates from 2012 and 2013 for adult samples at both Hastings and Soundview (both rivers), and for spat at Hastings (Hudson) demonstrate the temporal consistency of this river-specific pattern of admixture.

***AMOVA***

Temporal structure, hypothesized to be driven by sweepstakes reproductive success, was tested by using the Hastings on Hudson and Soundview population samples where we had both spat and adults sampled in both 2012 and 2013. Haplotype data from the 17 wild sample set including 4,186 loci were used. At both Hastings and Soundview, the AMOVA (pair-wise distance model) indicated there was no significant difference between samples in 2012 and 2013 (Phi_CT=0.001, proportion of variation=0.002, $P>0.05$, Table IV a), but there was a significant difference between spat and adults within year (HH Phi_ST= 0.339, proportion of variation=0.488, $P<0.001$; SV Phi_ST= 0.398, proportion of variance = 0.554, $P<0.001$). An AMOVA used to test between Hudson River and East River samples showed negligible regional differences between two-year classes (Table IV a). When the Hudson River and East River samples were aggregated to represent 2 wild populations for comparison to the two aquaculture samples, the between-group Phi_CT was relatively high when aquaculture strain samples (17+2) were included, showing substantial genetic differentiation between the wild and aquaculture oysters (Phi_CT= 0.111, proportion of variation=0.339, $P<0.01$). In contrast, there was no significant difference between populations within these two groups (Phi_ST=0.015, proportion of variation=0.001, $P>0.05$) (Table IV b).

***Functional Analysis of Outliers***

When the full wild population set of 4,186 loci were blasted to a *C. virginica* transcriptome the proportion with e-value hits $< 10^{-6}$ was 39.5%. Using the same procedures with the 51 outlier loci found that 19 loci matched transcriptome contigs (37.2%), close to what would

be expected for a random sample of our ddRAD loci. Annotations for the 51 outlier loci, mostly derived from the *C. gigas* reference genome (Eierman and Hare 2014) are given in Supplemental Table1.

## DISCUSSION

Due to densely populated coastlines in New York City and heavy shipping traffic, the Hudson River Estuary used to be considered one of the most polluted, nitrogen loaded estuaries in the world (Howarth *et al* 2006). This estuary still represents an extreme example of a severely depleted oyster population, functionally extirpated but for occasional sparse spat recruitment and one known remnant population in the low salinity portion of the Hudson (Medley 2010). Under these evolutionarily non-equilibrium conditions we expected that genetic population structure could inform us about contemporary or recent processes. Relevant processes could include selection, given the potentially strong selective pressures experienced at low salinity by the relatively isolated remnant oyster population, sweepstakes reproduction, or admixture with domesticated aquaculture strains. To efficiently build a large multilocus dataset we used ddRAD-Seq and tested for temporal and spatial patterns of differentiation within the Hudson-Raritan Estuary.

Similar to previous population genetic studies of this species, genetic variation was abundant and Hardy-Weinberg deviations were largely attributable to null alleles, a technical artifact that was easily filtered down to negligible levels by eliminating loci with missing data. Temporal and spatial genetic homogeneity was the predominant pattern observed among both spat and adults, contrary to expectations from sweepstakes reproduction but consistent with previous tests using microsatellites to examine single estuary populations of *C. virginica* (Rose *et al* 1996; He *et al* 2012). Although a model-based test for non-neutral patterns of elevated allele frequency differentiation identified 51 loci, their collective pattern of differentiation was not spatially consistent with the most obvious gradients in physical environmental conditions. In particular, the southern-most Hudson River sample, Pralls Island on the West side of Staten Island with moderate salinity, showed more similar genetic patterns to low-salinity Hudson River samples to the north versus East River samples. Instead, subtle differences in admixture patterns distinguished oysters in the Hudson and East Rivers, implying a source of introgression proximal to the East River and low connectivity between these rivers. Inclusion of an aquaculture reference sample in assignment test analyses showed spatially discreet patterns of admixture that suggest the East River oysters were subject to historical introgression from aquaculture strains,

presumably in the Long Island Sound populations that likely generate East River recruits through westward larval dispersal. This inference is strengthened by our replication across two years and conservative quality-control filtering of the ddRAD data. However, because we did not sample oysters in Long Island Sound and included only one representative aquaculture strain, this interpretation requires further study to confirm hybridization as the mechanism, identify the timing and source of introgression, as well as determine the functional consequences of this introgression.

***Genetic Homogeneity was the Predominant Pattern in the Wild***

The Structure analysis indicated that Hudson and East River oysters share two sources of ancestry, i.e. they are both admixed. There was no population structure in terms of the admixture level between the Hudson and East Rivers when all loci (4,184) were examined in the 17-wild population dataset, but in the absence of reference samples from relevant source populations, lack of spatial heterogeneity may have been due to low resolving power.

When high-differentiation outlier loci were identified and examined with assignment-based clustering, there was a subtle difference in the level of admixture between the Hudson and East Rivers. If the outlier loci were subject to selection along the Hudson River salinity gradient then we expected the more northern Hudson populations (#1 – 7 in Table I, all experiencing low salinity) to show discreet allele frequencies compared with the lower Hudson (Pralls Island) and East River populations experiencing moderate salinities. Instead, the major pattern of genetic differentiation separated samples from the Hudson and East Rivers, suggesting enough restriction in gene flow between the two rivers that a recent change in one would not immediately be reflected in the other. The restriction in the gene flow between the two rivers may be because the channel between the two rivers is very narrow. The narrowness can affect the direction of currents and advection, creating a barrier to complete larval mixing. To gain a deeper understanding of how this subtle separation between the two rivers can influence larval transport, the hydrodynamic features of the Hudson River Estuary should be explored in this context in future studies.

***Testing the Sweepstake Reproductive Success Hypothesis***

Another possible mechanism that can create spatial and temporal genetic differentiation with broadcast-spawners like eastern oysters is the sweepstake reproductive success hypothesis. If this is the case, then we expected to see some degree of difference in the genetic diversity of spat relative to adults. However, there was no detectable reduction in genetic diversity in terms of allelic richness or observed heterozygosity between these two age classes. In addition, there was no significant year-to-year difference when tested in adults and spat whereas SRS predicts heterogeneity across spat cohorts. The SRS also predicts genetic differences between adults (assumed to be mixed-cohort) and a single cohort of spat, and in two HRE locations there were genetic differences between adult samples and spat samples as indicated by substantial Phi_ST values. This pattern supports some variation in reproduction success, but is difficult to understand when corresponding predictions for the other contrasts were not born out. Interestingly, He *et al* (2012) had similar observations in examining the eastern oysters in Delaware Bay with seven nuclear microsatellites markers, citing no evidence for a cohort-effect in terms of allelic richness or observed heterozygosity over three different year classes, but finding the greatest support for genetic differentiation when comparing spat and adults. Thus, the results suggest that while some mechanism may involve a small set of parents and elevate genetic drift to some degree, such variance in reproductive success does not have a strong effect on the genetic variation of the eastern oysters in Hudson River Estuary.

***Detection and Influence of Selection***

The significant advantage to ddRADseq-based genome scans over traditional markers in marine population genetic studies is that the dramatic increase in number of markers increases the power of detecting subtle population structure caused by demographic mechanisms, and also increases the odds of sampling genetic variation subject to selection. Nonetheless, ddRAD sampling is typically sparse, in this study representing ~0.07% of the whole genome, so that in species with low linkage disequilibrium there is a very low probability of finding a ddRAD marker linked to a locus under selection (Lowry *et al* 2016). Indeed, the 51 high-Fst outliers identified with LOSITAN based on wild population samples in this study are more parsimoniously explained by spatially discreet patterns of introgression rather than selection. Although introgression is expected to affect the whole genome, rather than be locus specific, loci

that happen to differ more between the two source populations will show the biggest difference between introgressed and non-introgressed populations. Selection seems less parsimonious for two reasons. First, the Structure barplot based on these 51 loci showed a pattern consistent with East River admixture and without any obvious correlation to physical environmental gradients such as salinity. Second, the frequency of coding sequence among these 51 outlier loci was not exceptional relative to that for all the ddRAD loci from wild population samples, rejecting a hypothesized enrichment if coding sequence variation was in fact the target of selection. Surprisingly, 40% of all our ddRAD loci matched coding sequences in the *C. virginica* transcriptome, whereas only 2% was expected given the proportion of the *C.gigas* genome that is coding (Zhang *et al* 2012; Eierman and Hare 2015). It appears that the ddRAD procedures used here preferentially sampled from transcribed sequences. The restriction sites for the two ddRAD enzymes, PstI (CTGCAG) and NlaIII (CATG), had a GC bias, potentially creating biased sampling of coding regions over non-coding regions. An additional potential source of bias could result from the ddRAD locus filtering. If ddRAD restriction sites in coding sequences had a lower likelihood of being polymorphic, compared to noncoding sites, then our filtering of loci with null alleles would have removed more noncoding loci relative to coding loci.

***Characteristics of Aquaculture Strain***

Two population samples of the Fisher's Island aquaculture strain were included in the clustering analysis because recent oyster restoration efforts (ORRP) in both the Hudson and East rivers have included plantings of this strain (Grizzle *et al* 2013). Our expectation was that new spat recruits might have Fisher's Island parentage, or have aquaculture x wild parentage from interbreeding, and the location of these recruits could be informative about larval dispersal from the spatially discreet restoration sites.

For the 3,655 ddRAD loci that provided high quality comparability among wild population samples and the Fisher's Island strain, population genomic analyses provided two insights. The first finding was that the aquaculture strain had substantially lower genetic variation than the wild, as expected based on similar studies conducted on cultivated Atlantic salmon and Pacific abalone strains when these were compared to wild progenitor stocks (Hedgecock 1990; Clifford *et al* 1998; Li *et al* 2004; Yu and Guo 2005; Zhong *et al* 2016). This reduced genetic diversity in aquaculture strains may not be a limiting factor within the

aquaculture industry if inbreeding is minimized, but it suggests caution is warranted with respect to the use of aquaculture strains for restoration and population supplementation.

In a restoration context, low genetic variation is a concern because genetic variation is known to be important for long-term persistence of populations given that the level of additive genetic variation determines their adaptability to environmental changes (Fisher 1958; Li 2007). Yu and Guo (2005) detected that the number of rare alleles of the 4 selected strains decreased significantly compared with a wild population in Eastern oyster. The reduction of genetic diversity in a population can reduce disease resistance and decrease adaptability to environmental changes (Yu and Guo 2005). Thus, although there is recent use of domesticated oyster aquaculture strains (Hare *et al* 2006; Carlsson *et al* 2008) for population supplementation, this raises questions about the impact such activities can have on the conservation and restoration of natural populations ( Roodt-Wilding 2007; Bester-van *et al* 2011).

The second finding was that the aquaculture strain showed distinctive allele frequencies when compared to the wild populations, providing the analytical power to detect interbreeding and introgression. Although we were able to include only one aquaculture strain in this study, it is expected that any product of domestication and selective breeding will show differences resulting from both genetic drift and artificial selection. Genetic differences from drift would be expected throughout the genome whereas those generated by artificial selection are likely to be localized near loci under selection.

***Implications of Introgression from the Aquaculture Strain***

Contrary to our expectations we did not see localized evidence of very recent Fishers Island strain reproduction, or interbreeding with wild oysters at Hastings and Soundview. Instead, patterns of admixture were primarily restricted to the East River and inferred from two distinct patterns in the STRUCTURE bar plots. First, the admixture attributed to Fisher's Island strain oysters was observed in every East River oyster sampled, including spat and adults, consisting of Q values as high as 0.318 and with a mean and SD of 0.107 +/- 0.11(95% CI). This low level of admixture, coupled with the spatial uniformity of its occurrence within the East River, implies that it derives from introgression and backcrossing over several generations in the past, possibly from interbreeding in Long Island Sound rather than in the Hudson River Estuary. The other genetic indication of admixture in the K=3 result involves two source populations that contribute

to both Hudson and East River samples, but a high variance in Q values were inferred among East River samples, with mean 0.063 +/- 0.047 (95% CI), compared with no admixture in a majority of Hudson River individuals and where present, Q values averaging 0.002 +/- 0.018 (95% CI). Because Fisher's Island is only one of many aquaculture strains that are planted on oyster farm lease sites in Long Island Sound, and aquaculture strains have independent histories of genetic drift and artificial selection leading to genetic differentiation, it is possible that this subtle signal of admixture in the East River could stem from other aquaculture strains. We predict that inclusion of reference samples from additional aquaculture strains would parse some of this general admixture background into strain-specific components. It is possible that all aquaculture strains experience parallel genomic changes in response to domestication selection and these differences from wild oysters would not be attributable to any one strain, but would continue to appear as background admixture.

Without additional aquaculture strain reference samples, it is difficult to interpret these patterns with complete confidence. However, one possibility is that all aquaculture strains evolve parallel changes in response to domestication (tank performance as larvae, spawning predictability, fast growth) so that any strain can be recognized as the source of introgression even though interbreeding between wild and aquaculture oysters could involve many strains of the latter. This hypothesis requires comparisons among aquaculture strains and inclusion of additional reference samples in assignment test admixture analyses of wild populations.

***Implication for the Success of Oyster Restoration***

In the Hudson River Estuary, improved water quality has motivated the large-scale restoration of oyster populations. Wild recruitment from surrounding populations may be an important source of population recovery in waters near New York City. The East River, in particular, has no self-sustaining adult breeding population, or a sparse one at best. Adult samples analyzed here were survivors from a recent cohort of recruits and did not persist (Hare personal observation). This study showed that historical gene flow from aquaculture strains into wild oyster populations has uniformly affected the genetics of recruiting spat in the East River for two consecutive years. These admixed recruits indicate that the source population of admixed oyster stocks produce viable offspring. On the other hand, the low level of aquaculture introgression found in East River oysters could be due to strong selection against greater degrees

of introgression. Finally, it is also possible to explain the patterns without assuming differential fitness of hybrid progeny given that the low level of introgression is consistent with hybridization several generations back with subsequent back crossing of hybrid offspring to wild oysters.

The fact that small amounts of admixture are evident in the Hudson River indicates that gene flow is not entirely absent between the two rivers, but over the historical timescale of introgression in Long Island Sound populations, larval dispersal mediated gene flow of aquaculture genes into Hudson River populations has been limited. If the start of significant introgression can be dated for Long Island Sound, or temporal replication can determine the influx rate of aquaculture genes into the Hudson, then the rate of gene flow from East to Hudson Rivers can be measured in the future.

***Limitations***

Although genetic methods in general and RADseq in particular offer great potential for measuring marine population demography (Gagnaire *et al* 2015), there were several limitations and caveats with respect to this study. First, observed patterns suggested mechanistic hypotheses that are impossible to test without additional sampling of natural stocks in Long Island Sound and additional domesticated strains of aquaculture oyster. Secondly, large amounts of missing genotype data due to null alleles required large numbers of loci to be filtered out. This procedure only minimized null allele effects, rather than eliminate them, and severely reduced our genomic sampling. Nonetheless, our procedures provided for robust analyses of a large sample of loci in this highly heterozygous species. Tests for genetic – environmental correlations may be subject to false positives because of the gapped spatial arrangement of population samples along the salinity gradient. Specifically, by leaving a spatial sampling gap across a large portion of the moderate salinity gradient we strengthened patterns of autocorrelation that can undermine tests of selection.

## CONCLUSIONS and RECOMMENDATIONS

Our results showed that a subtle pattern of genomic population structure existed in one portion of the Hudson River Estuary despite the high dispersal potential of oyster larvae. This spatial heterogeneity is most strongly manifest in the highest Fst loci, but is a very general pattern affecting many loci. Our tentative interpretation is that this structure derives from recent admixture between domesticated aquaculture strains and wild oysters. This study did not evaluate fitness-related performance so the functional consequences of introgression from aquaculture oysters are unknown. Until fitness consequences of introgression, and relative fitness of wild and aquaculture oysters can be evaluated over the full oyster life cycle, we recommend stricter adherence to published recommendations (Bain *et al* 2007; The Nature Conservancy 2012) that wild broodstock be used for generating oyster restoration seed, and to avoid wild/aquaculture plantings in close proximity. Oyster aquaculturists in other regions have had success with triploid oysters because they are nonreproductive and therefore devote more energy to somatic growth (Suquet *et al* 2016). Should these patterns of genetic admixture prove to be temporally and spatially more extensive, facilitating easier access to hardy triploid aquaculture seed could provide a win-win solution.


### *Acknowledgements*

This research was funded by NY Sea Grant (NYSG R/XG-22 ). We are greatly appreciative of a scholar fellowship provided to YK by NY Sea Grant and the Woodswoman Scholarship Fund provided by Department of Natural Resources at Cornell University. We thank Jeff Levinton, Pete Malinowski and Fred Jacobs for generously providing some samples analyzed here. Thanks also to Jim Lodge and Meredith Comi for assistance with field logistics. We are grateful to Dan Illut, Mark Ravinet and the Cornell BioHPC for help with bioinformatics, and Andrew Bange for primary QC filtering scripts. We are extremely grateful to members of the Hare Lab including support in the field and lab by Gavriela Carver, Jake Leiby, Katie McFarland, Harmony Borchardt-Wier, Catherine Sun, and Jennifer Zhu.

## REFERENCES


Allen Jr., S. K., P. M. Gaffney, and J. W. Ewart. (1993). Genetic Improvement of the Eastern Oyster for Growth and Disease Resistance in the Northeast in N. R. A. Center. Dartmouth, Massachusetts.

Antao, T., Lopes, A., Lopes, R. J., Beja-Pereira, A., & Luikart, G. (2008). LOSITAN: A workbench to detect molecular adaptation based on a F st -outlier method. *BMC Bioinformatics*, *9*(1), 1–5. http://doi.org/10.1186/1471-2105-9-323

Araki, H., Cooper, B., & Blouin, M. S. (2017). Linked references are available on JSTOR for this article : Genetic Effects of Captive Breeding Cause a Rapid , Cumulative Fitness Decline in the Wild, *318*(5847), 100–103.

Bain M, Lodge J, Suszkowski D, Matuszeski W (2007) Target Ecosystem Characteristics for the Hudson Raritan Estuary: A technical guide for developing a conprehensive restoration plan.

Benestan L, Gosselin T & Perrier C. (2015) RAD-genotyping reveals fine-scale genetic structuring and provides powerful population assignment in a widely distributed marine spe- cies; the American lobster (Homarus americanus). Molecular Ecology

Bester-van der Merwe, A. E., Roodt-Wilding, R., Volckaert, F. A. M., & D'Amato, M. E. (2011). Historical isolation and hydrodynamically constrained gene flow in declining populations of the South-African abalone, Haliotis midae. *Conservation Genetics*, *12*(2), 543–555. http://doi.org/10.1007/s10592-010-0162-0

Billion Oyster Project. 2013.Retrieved from https://www.billionoysterproject.org

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J. H., & De Koning, D.-J. (2011). Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes|Genomes|Genetics*, *1*(3), 171–182. http://doi.org/10.1534/g3.111.000240

Catchen, J. M. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140. http://doi.org/10.1111/mec.12354.Stacks

Carlsson J., Carnegie B. R., Cordes F. J., Hare P.M., Leggett T.A., Reece S.K. (2008) Evaluating recruitment contribution of a selectivity based aquaculture line of the oyster, *Crassostrea virginica* used for restoration efforts. J. Shellfish Res., 27 pp1117-1124.

Christie MR, Johnson DW, Stallings CD, & Hixon MA (2010) Self-recruitment and sweepstakes reproduction amid exten-sive gene flow in a coral-reef fish. Molecular Ecology, 19, 1042-1057.

Christie MR, MarineML, French RA, Waples RS & BlouinMS (2012) Effective size of a wild salmonid population is greatly reduced by hatchery supplementation. Heredity 109: 254−260.

Clifford, S. L., P. McGinnity, and A. Ferguson. 1998. Genetic changes in Atlantic salmon (Salmo salar) populations of northwest Irish rivers resulting from escapes of adult farm salmon. Can. J. Fish. Aquat.Sci. 55:358-363.

Cockerham, CC and Weir BS. (1993). Estimation of gene flow from F-statistics. Evolution 4:855-863.

Coen, L. D., Brumbaugh, R. D., Bushek, D., Grizzle, R., Luckenbach, M. W., Posey, M. H., … Tolley, S. G. (2007). Ecosystem services related to oyster restoration. *Marine Ecology Progress Series*, *341*, 303–307. http://doi.org/10.3354/meps341299

Corlett, R. T. (2015). The Anthropocene concept in ecology and conservation. *Trends in Ecology & Evolution*, *30*(1), 36–41. http://doi.org/10.1016/j.tree.2014.10.007

Earl DA, von HoldtBM(2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359-361.

Eierman, L. E., & Hare, M. P. (2014). Transcriptomic analysis of candidate osmoregulatory genes in the eastern oyster *Crassostrea virginica*. *BMC Genomics*, *15*(1), 503. http://doi.org/10.1186/1471-2164-15-503

Eierman LE, Hare MP (2016). Reef-specific patterns of gene expression plasticity in Eastern oysters (*Crassostrea virginica*) J. Hered.107(1):90–100. doi: 10.1093/jhered/esv057.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA & Kawamoto K. (2011). A robust, simple - Genotyping-by-Sequencing (GBS) approach for high diversity species. PLOS ONE 6.

Evanno G, Regnaut S, Goudet J (2005) Detecting the num-ber of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14: 2611-2620

Frank-Lawale, A., Allen, S. K., & Dégremont, L. (2014). Breeding and Domestication of Eastern Oyster ( *Crassostrea virginica* ) Lines for Culture in the Mid-Atlantic, Usa: Line

Development and Mass Selection for Disease Resistance. *Journal of Shellfish Research*, *33*(1), 153–165. http://doi.org/10.2983/035.033.0115

Fisher, R.A. (1958). The genetical theory of natural selection. Dover, New York.

GaffneyP.M. (1996). Biochemical and population genetics, in The eastern oyster *Crassostrea virginica*, V. S. Kennedy, R. I. E. Newell and A. F. Eble, eds., College Park: Maryland Sea Grant College, 423–441.

Goulletquer, P., Gros, P., Boeuf, G., & Weber, J. (2014). Biodiversity in the marine environment. *Biodiversity in the Marine Environment*, *9789401785*(Palumbi 2001), 1–198. http://doi.org/10.1007/978-94-017-8566-2_1

Grizzle, Raymond ; Ward, Krystin; Lodge, Jim; Suszkowski, Dennis; Mosher-Smith, Katie; Kalchmayr, Kerstin; Malinowski, P. (2013). Oyster restoration research p roject (ORRP) final technical report ORRP Phase I : Experimental Oyster Reef Development and.

Grizzle, R. E., & Brumbaugh, R. D. (2013). Quantifying the Loss of a Marine Ecosystem Service : Filtration by the Eastern Oyster in US Estuaries, 36–43. http://doi.org/10.1007/s12237-012-9559-y

He, Y., Ford, S. E., Bushek, D., Powell, E. N., Bao, Z., & Guo, X. (2012). Effective population sizes of eastern oyster <*Crassostrea virginica*> (Gmelin) populations in Delaware Bay, USA. *Journal of Marine Research*, *70*(2), 357–379. http://doi.org/10.1357/002224012802851977

Hedgecock, D. (1986). Is gene flow from pelagic larval dispersal important in the adaptation and evolution of marine invertebrates? . *Bull.Mar.Sci.* , *39*(2), 550–564.

Hedgecock, D. & Sly, F. (1990). Genetic drift and effective population sizes of hatchery-propagated stocks of the Pacific oyster *Crassostrea gigas*. Aquaculture 88, 21–38.

Hedgecock, D. (1994). Does variance in reproductive success limit effective population size in marine organisms? *Genetics and Evolution of Aquatic Organisms*, (January 1994), 122–134.

Hedgecock, D., & Pudovkin, A. I. (2011). SWEEPSTAKES REPRODUCTIVE SUCCESS IN HIGHLY FECUND MARINE FISH AND SHELLFISH : A REVIEW AND COMMENTARY Dennis Hedgecock and Alexander I Pudovkin. *Bulletin of Marine Science*, *87*(4), 971–1002. http://doi.org/10.5343/bms.2010.1051

Hedrick, P. (2005). Large Variance in Reproductive Success and the Ne / N Ratio Author ( s ):

Philip Hedrick Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org/stable/3449182 JSTOR is a not-for-profit service that helps scholars , researchers, *59*(7), 1596–1599.

Howarth, R. W., & Marino, R. (2006). Nitrogen as the limiting nutrient for eutrophication in coastal marine ecosystems : Evolving views over three decades, *51*, 364–376.

Hubisz M, Falush D, Stephens M, Pritchard JK. Inferring weak population structure with the assistance of sample group information. Molecular Ecology Resources. 2009; 9:1322–1332.

Ilut DC, Nydam ML & Hare MP (2014) Defining loci in restric-tion-based reduced representation genomic data from non-model species: sources of bias and diagnostics for optimal clustering. BioMed Research International.

Jackson, J. B. C., Kirby, M. X., Berger, W. H., Bjorndal, K. A., Botsford, L. W., Bourque, B. J.,Warner, R. R. (2001). Historical Collapse Overfishing of and the Recent Coastal Ecosystems.

Jombart T., Devillard S., Balloux F. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC genetics* 11:94.

Johnson, Michael; Black, R. (1984). Pattern Beneath the Chaos : The Effect of Recruitment on Genetic Patchiness in an Intertidal Limpet Author ( s ): Michael S . Johnson and Robert Black Reviewed work ( s ): Published by : Society for the Study of Evolution Stable URL : http://www.jstor.org. *Society for Study of Evolution*, *38*(6), 1371–1383.

Kareiva, P. (2014). New Conservation : Setting the Record Straight and Finding Common Ground, *28*(3), 634–636. http://doi.org/10.1111/cobi.12295

Karl S. A. and J. C. Avise. (1992). Balancing selection at allozyme loci in oysters: implications from nuclear RFLPs. Science, 256, 100–102.

Kennedy VS, 1996. The biology of larvae and spat. In: The eastern oyster: *Crassostrea virginica* (Kennedy VS, Newell RIE, and Eble AF, eds). College Park, MD: Maryland Sea Grant; 371–422.

Kennish M.J. 1992. Ecology of Estuaries: Anthropogenic Effects. CRC Press, Boca Raton, FL, 494 pp.

Lennon, M. (2017). Nature conservation in the Anthropocene : preservation , restoration and the challenge of novel ecosystems, *9357*(May). http://doi.org/10.1080/14649357.2015.1027047

Li, Q., Park, C., Endo, T., Kijima, A. (2004). Loss of genetic variation at microsatellite loci in hatchery strains of the Pacific abalone (Haliotis discus hannai). Aquaculture 235, 207–222.

Li, Y., J. G. Qin, C. A. Abbott, X. Li & K. Benkendorff. (2007). Synergistic impacts of heat shock and spawning on the physiology and immune health of *Crassostrea gigas*: an explanation for sum-mermortality in Pacific oysters. Am. J. Physiol. Reg. I 293:R2353–R2362.

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF and Storfer A.(2016) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Mol. Ecol. Resour. 2017;17:142–152. doi: 10.1111/1755-0998.12635.

MacKenzie, C.L. (1989). A guide for enhancing estuarine molluscan shellfisheries. Marine Fisheries Review 51(3): 1-47.

MacKenzie, C.L. (1996). History of oystering in the United States and Canada, featuring North America's greatest oyster estuaries. Marine Fisheries Review 58(4): 1-78.

Mann, R., & Powell, E. N. (2007). Why Oyster Restoration Goals in the Chesapeake Bay Are Not and Probably Cannot Be Achieved. *Journal of Shellfish Research*, *26*(4), 905–917. http://doi.org/10.2983/0730-8000(2007)26[905:WORGIT]2.0.CO;2

Hare, M.P., Allen, S.K., Bloomer, P., Camara, M.D., Carnegie, R.B., Murfree, J., Luckenbach, M., Meritt, D., Morrison, C., Paynter, K. and Reece, K.S., 2006. A genetic test for recruitment enhancement in Chesapeake Bay oysters, Crassostrea virginica, after population supplementation with a disease tolerant strain. *Conservation Genetics*, *7*(5), pp.717-734.

McCay, B.J. (1998). Oyster Wars and the Public Trust: Property, Lawn, and Ecology in New Jersey History. University of Arizona Press, Tucson, 246pp.

Medley, T. L. (2010). Wild oysters, *Crassostrea virginica*, in the Hudson River Estuary: Growth, health and population structure. Ph.D. Dissertation. City University of New York. pp1-147.

Miller, D. E. (2013). Hudson River Estuary Habitat Restoration Plan. New York State Department of Environmental Conservation, Hudson River Estuary Program.

Murcia, C., Aronson, J., Kattan, G. H., Moreno-mateos, D., Dixon, K., & Simberloff, D. (2014). A critique of the " novel ecosystem " concept. *Trends in Ecology & Evolution*, *29*(10), 548–553. http://doi.org/10.1016/j.tree.2014.07.006

Narváez, D.A., Klinck, J.M., Powell, E.N., Hofmann, E.E., Wilkin, J., & Haidvogel, D.B. (2012). Modeling the dispersal of eastern oyster (*Crassostrea virginica*) larvae in Delaware Bay. Journal of Marine Research, 70(2-3), 381-409.

National Marine Fisheries Services (2016). Fisheries of the United States, 2015. U.S. Department of Commerce, NOAA Current Fishery Statistics No.2015. Available at: https://www.st.nmfs.noaa.gov/commercial- sheries/ fus/fus15/index

North, E.W., Schlag, Z., Hood, R.R., Li, M., Zhong, L., Gross, T., Kennedy, V.S. (2008). Vertical swimming behavior influences the dispersal of simulated oyster larvae in a coupled particle-tracking and hydrodynamic model of Chesapeake Bay. Marine Ecology Progress Series 359, 99e115.

NYC DCP. 2013. Coastal Climate Resilience: Urban Waterfront Adaptive Strategies. Collaboration of The City of New York and Department of City Planning.

Palumbi, S. R., & Mu, P. (2001). Humans as the World's Greatest Evolutionary Force The Pace of Human-Induced Evolution, *293*(September), 1786–1790.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double Digest RADseq: An Inexpensive Method for *De Novo* SNP Discovery and Genotyping in Model and Non-Model Species. PLoS ONE 7(5): e37135. doi:10.1371/journal.pone.

Pomeroy, L.R., C.F. D'Elia, and L.C. Schaffner. (2006). Limits to top-down control of phytoplankton by oysters in Chesapeake Bay. Marine Ecology Progress Series 325: 301–309.

Pritchard JK, Stephens M & Donnelly P (2000). Inference of population structure using multilocus genotype data. Genetics. 155:945–959.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reeb, C. A., & Avise, J. C. (1990). A genetic discontinuity in a continuously distributed species: Mitochondrial DNA in the American oyster, *Crassostrea virginica*. *Genetics*, *124*(2), 397–406.

Roodt-Wilding, R. (2007). Abalone ranching: A review on genetic considerations. *Aquaculture*

*Research*, *38*(12), 1229–1241. http://doi.org/10.1111/j.1365-2109.2007.01801.x

Rose, C. G., Paynter, K. T., & Hare, M. P. (2006). Isolation by distance in the eastern oyster, *Crassostrea virginica*, in Chesapeake Bay. *Journal of Heredity*, *97*(2), 158–170. http://doi.org/10.1093/jhered/esj019

Ryman N, Utter F, and Laikre L. (1995). Protection of intraspecific biodiversity of exploited fishes. Rev Fish Biol Fish 5:417–446.

Soria-Carrasco V, Gompert Z, Comeault AA, Farkas TE, Parchman TL, Johnston JS, Buerkle CA, Feder JL, Bast J, Schwander T. (2014). Stick insect genomes reveal natural selection's role in parallel speciation. Science. 344:738–742.

Suquet M., Malo F., Quere C., Ledu C., Le Grand J. (2016). Gamete quality in triploid pacific oyster (crassostrea gigas). *Aquaculture* 451:11–15.

Stanne S. (2015).The State of the Hudon 2015. Hudson River Estuary Program. New York State Department of Environmental Conservation.

Stenzel, H.B. (1971). Oysters. Treatise on Invertebrate Paleontology. Part N, Bivalvia 3:953-1224.

Taris, N., Batista, F. M., & Boudry, P. (2007). Evidence of response to unintentional selection for faster development and inbreeding depression in *Crassostrea gigas* larvae. *Aquaculture*, *272*(SUPPL. 1), 69–79. http://doi.org/10.1016/j.aquaculture.2007.08.010

Thompson, R.J., R.I.E. Newell, V.S. Kennedy and R. Mann. (1996). Reproductive processes and early development.The Eastern Oyster *Crassostrea virginica*. Maryland Sea Grant College, University of Maryland, College Park, Maryland. pp. 335-370.

Turner, T. F., Wares, J. P., & Gold, J. R. (2002). Genetic effective size is three orders of magnitude smaller than adult census size in an abundant, estuarine-dependent marine fish (Sciaenops ocellatus). *Genetics*, *162*(3), 1329–1339.

Waldman, J. (1999). Heartbeats in the Muck. The Lyons Press, NY.

White, B. (2013) Genotyping by Sequencing (ddRadSeq) Protocol for Anopheles. https://static1.squarespace.com/static/5154a345e4b0feb1eb1819fc/t/52a7681ce4b0253945d2e24d/1386702876160/White+Lab+GBS+ddRAD.pdf

Worm, B., Barbier, E. B., Beaumont, N., Duffy, J. E., Folke, C., Halpern, B. S., … Watson, R. (2006). Impacts of biodiversity loss on ocean ecosystem services. *Science (New York, N.Y.)*, *314*, pp787–790.

Wright, S. (1931). Evolution in Mendelian populations. Genetics 16:114-138

Yozzo, D. J., Wilber, P., & Will, R. J. (2004). Beneficial use of dredged material for habitat creation , enhancement , and restoration in New York – New Jersey Harbor, *73*, 39–52. http://doi.org/10.1016/j.jenvman.2004.05.008

Yu ZN & Guo XM. (2005). Genetic analysis of selected strains of Eastern oyster (*Crassostrea virginica* Gmelin) using AFLP and microsatellite markers. Mar Biotechnol. 6: 575–586.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H, Xiong Z, Que H, Xie Y, Holland PWH, Paps J, Zhu Y, Wu F, Chen Y, Wang J, Peng C, Meng J, Yang L, Liu J, Wen B, Zhang N, Huang Z, Zhu Q, Feng Y, Mount A, and Hedgecock D. (2012). The oyster genome reveals stress adaptation and complexity of shell formation. Nature 490:49–54.

## Figures

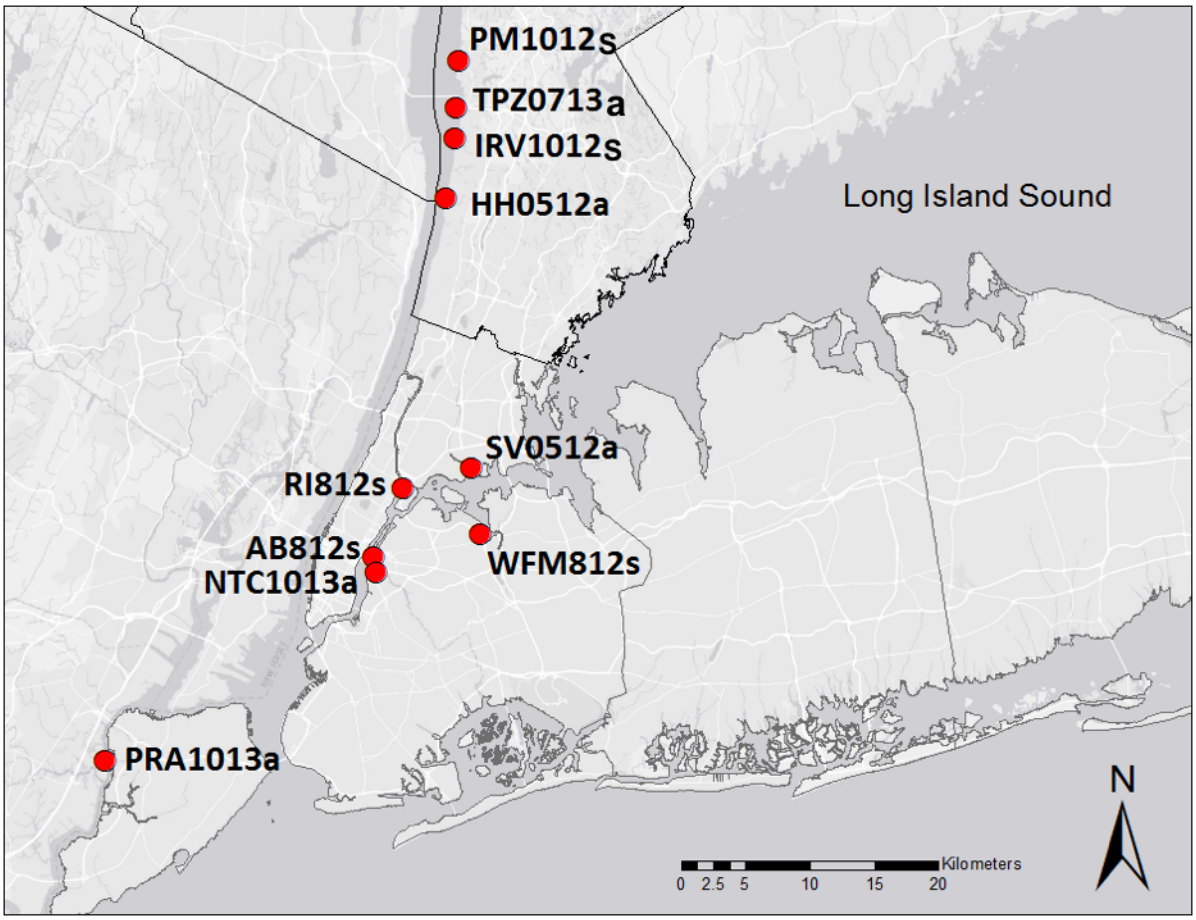
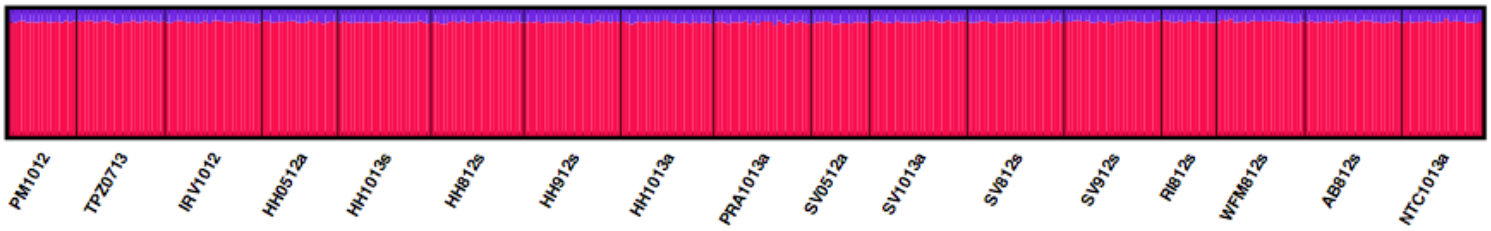
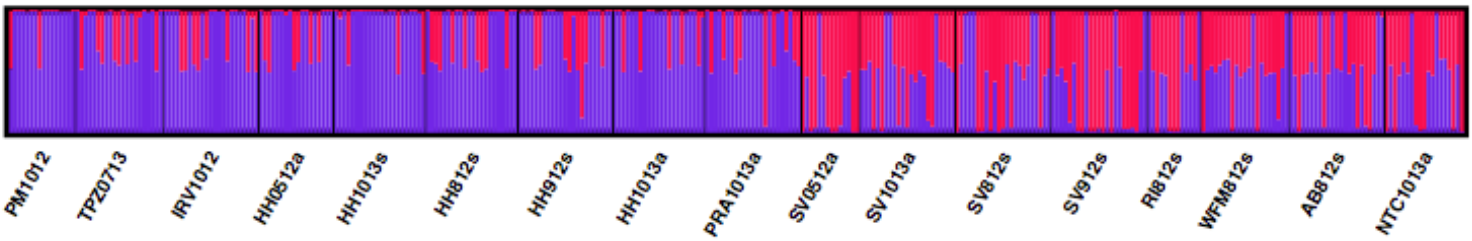


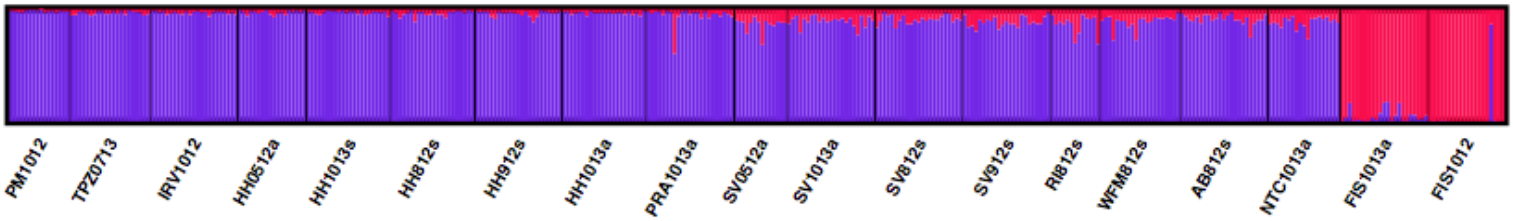

**Figure 1:** Map of the eastern oyster *(C. virginica)* collection sites from the Hudson-Raritan Estuary, NY. The location abbreviations include site code and month-year of collection as listed in Table 1. Twenty-four oysters, either adults or spat ("a" or "s" in the locality codes), were collected from each site in 2012 and 2013. HH and SV sites had collections of both spat and adults in both 2012 and 2013.



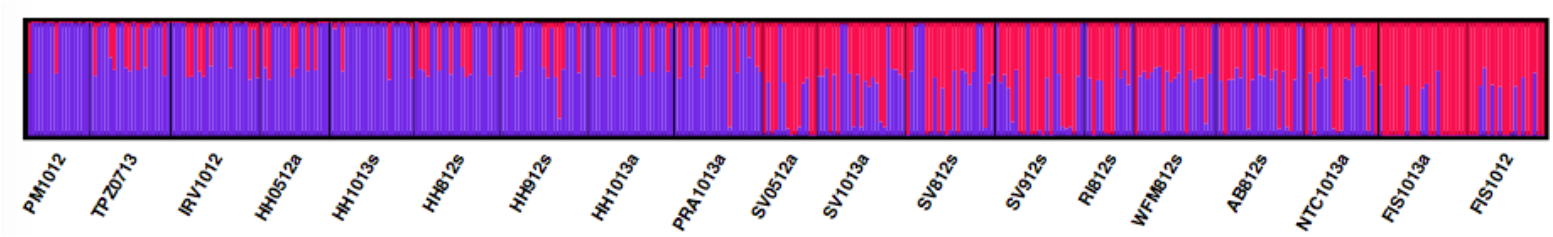

a)




b)




c)




d)

**Figure 2: Assignment of population clusters using haplotype data in STRUCTURE.**

In all cases the MCMC had 100,000 burn-in iterations followed by 500,000 iterations for three replicates at each K from K=1 to 5, conducted with and without locality prior. Results shown included the locality prior. With all 4184 ddRAD loci, all 17 wild samples showed similar admixture patterns for the best supported K=2 analysis (a). Restricting analysis to 51 outlier loci from 17-wild data showed different admixture levels between the Hudson River and East River

(b). The aquaculture strain reference samples (FIS1013a, FIS1012) were mostly non-admixed and contributed more ancestry to East River oysters than to Hudson River samples (c). The outlier loci (N=51) defined from the 17-wild data, when re-examined in the context of the 19-sample dataset using a white list, showed a striking difference in the admixture level between the Hudson River and East River (d).

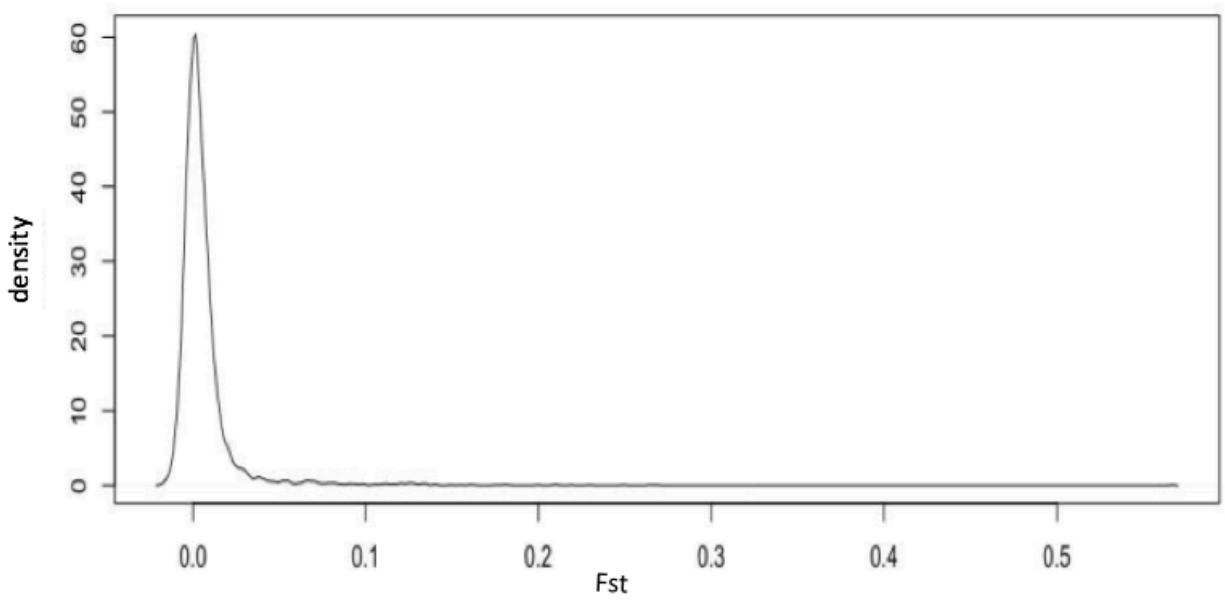



**Figure 3: The Fst Distribution of 4186 haplotype ddRAD loci (N=349)**

The Fst distribution for 17 population samples based on haplotype data at 4186 ddRAD loci. The median Fst was 0.002, and the mean was 0.006 with the range -0.001 to 0.566.

The Fst value was obtained from phistats.tsv outputfile from STACKS based on STACKS command-line filtering including—fstats. The Wier and Cockerham equation was used to calculate Fst.

## Tables

**Table I:** Sample names, locations, post-filtering sample sizes and sequencing read statistics. The "a" refers to adults in the sample name and the "s" refers to spats. Raw read counts, post *process_radtags*, and genomic averages after trimming to 89bp length were also included.

| Site Index, Site Name | Site Code | Life History | Latitude | Longitude | N | Raw read count | Post *process_radtags* | Genomic Average |
|---|---|---|---|---|---|---|---|---|
| 1.Phillipse Manor | PM1012s | spats | 41 05.691 | -73 52.220 | 16 | 4585332 | 3197541 | 2418865 |
| 2.Tappan Zee Bridge | TPZ0713a | adults | 41 03.750 | -73 52.433 | 21 | 6140831 | 3965768 | 3009449 |
| 3.Irvington Boat Club | IRV1012s | spats | 41 02.483 | -73 52.517 | 23 | 4787730 | 3382521 | 2456393 |
| 4.Hastings on the Hudson | HH0512a | adults | 40 59.977 | -73 53.058 | 18 | 5388135 | 3740279 | 2904547 |
| 5.Hastings on the Hudson | HH0812s | spats | 40 59.977 | -73 53.058 | 22 | 3335510 | 1982386 | 1525866 |
| 6.Hastings on the Hudson | HH0912s | spats | 40 59.977 | -73 53.058 | 23 | 6553640 | 4765472 | 3784088 |
| 7.Hastings on the Hudson | HH1013a | adults | 40 59.977 | -73 53.058 | 22 | 4169900 | 3465395 | 2695838 |
| 8.Hastings on the Hudson | HH1013s | spats | 40 59.977 | -73 53.058 | 22 | 3528788 | 2798578 | 2200823 |
| 9.Pralls Island, Staten Island | PRA1013a | adults | 40 36.549 | -74 12.162 | 23 | 5811566 | 4104818 | 3009413 |
| 10.Soundview Park | SV0512a | spats | 40 48.624 | -73 51.857 | 14 | 4380170 | 2754158 | 2149783 |
| 11.Soundview Park | SV0812s | spats | 40 48.624 | -73 51.857 | 23 | 4020871 | 3237320 | 2503537 |
| 12.Soundview Park | SV0912s | adults | 40 48.624 | -73 51.857 | 23 | 3903677 | 3147672 | 2456366 |
| 13.Soundview Park | SV1013a | spats | 40 48.624 | -73 51.857 | 23 | 4449971 | 3519400 | 2720829 |
| 14.Randall's Island | RI812s | spats | 40 47.842 | -73 55.637 | 13 | 2468801 | 1582169 | 1239793 |
| 15.Worlds Fair Marina | WFM812s | spats | 40 45.852 | -73 51.391 | 21 | 6212544 | 4840795 | 3644872 |
| 16.Anable Basin | AB812s | adults | 40 44.957 | -73 57.288 | 23 | 5929110 | 4459731 | 3565733 |
| 17.Newtown Creek | NTC1013a | adults | 40 44.338 | -73 57.178 | 19 | 3487688 | 2164986 | 1603839 |
| 18.Fisher's Island 2012 | FIS1012a | adults | unknown | unknown | 21 | 3602488 | 2143007 | 1683705 |
| 19.Fisher's Island 2013 | FIS1013a | adults | unknown | unknown | 23 | 3485167 | 2144507 | 1647168 |

**Table II.** Number of *C.virginica* individuals sampled and number of ddRAD loci before and after quality control (QC) filtering based on 17 wild populations. The final data for 4186 ddRAD loci were output as haplotypes for most analyses.

| **Parameter** | **Value** |
|---|---|
| Individuals sequenced | 394 |
| Individuals with >0.5 M reads | 349 |
| Initial ddRAD tags | 60206 |
| ddRAD tags following STACKS RX (*rx_stacks*) | 9250 |
| Initial ddRAD locus count *(m6r98p1maf001)* | 4212 |
| ddRAD loci after Fis filtering ($-0.5<F_{is}<0.5$) | 4186 |
| Final ddRAD loci | 4186 |

**Table III. Summary statistics from Genalex**

Per–population number of alleles, number of private alleles and the expected heterozygosity ($H_e$) were calculated with 17+2 population sample data with 13 individuals randomly selected from each sample population (N=13). The mean and the standard errors of these data were calculated with the haplotype (A) and the single random SNP (B) data for mostly the same ddRAD loci.

a. Haplotype (4092 loci, N=13 per population)

| Pop ID | N | No. of Alleles (Mean) | No. of Alleles (SE) | No. Private Alleles (Mean) | No. Private Alleles (SE) | $H_e$ (Mean) | $H_e$ (SE) |
|---|---|---|---|---|---|---|---|
| ***Wild Strains*** | | | | | | | |
| *Hudson River* | | | | | | | |
| PM1012s | 13 | 3.172 | 0.036 | 0.308 | 0.010 | 0.365 | 0.005 |
| TPZ0713a | 13 | 2.989 | 0.035 | 0.314 | 0.010 | 0.351 | 0.005 |
| IRV1012s | 13 | 3.499 | 0.037 | 0.476 | 0.013 | 0.399 | 0.005 |
| HH0512a | 13 | 3.208 | 0.035 | 0.337 | 0.010 | 0.378 | 0.005 |
| HH0812s | 13 | 3.038 | 0.037 | 0.315 | 0.010 | 0.351 | 0.005 |
| HH0912s | 13 | 2.835 | 0.036 | 0.261 | 0.009 | 0.330 | 0.005 |
| HH1013a | 13 | 3.374 | 0.037 | 0.359 | 0.011 | 0.388 | 0.005 |
| HH1013s | 13 | 3.203 | 0.036 | 0.328 | 0.010 | 0.372 | 0.005 |
| PRA1013a | 13 | 3.082 | 0.036 | 0.279 | 0.009 | 0.357 | 0.005 |
| *East River* | | | | | | | |
| SV0512a | 13 | 3.334 | 0.036 | 0.304 | 0.010 | 0.386 | 0.005 |
| SV0812s | 13 | 3.172 | 0.037 | 0.323 | 0.010 | 0.368 | 0.005 |
| SV0912s | 13 | 3.333 | 0.036 | 0.348 | 0.011 | 0.382 | 0.005 |
| SV1013a | 13 | 2.928 | 0.035 | 0.252 | 0.009 | 0.344 | 0.005 |
| RI812s | 13 | 2.975 | 0.036 | 0.270 | 0.009 | 0.348 | 0.005 |
| WFM812s | 13 | 3.149 | 0.036 | 0.264 | 0.009 | 0.363 | 0.005 |
| AB812s | 13 | 3.161 | 0.036 | 0.272 | 0.009 | 0.365 | 0.005 |
| NTC1013a | 13 | 3.187 | 0.036 | 0.307 | 0.010 | 0.371 | 0.005 |
| ***Aquaculture*** | | | | | | | |
| FIS1013a | 13 | 2.124 | 0.023 | 0.080 | 0.005 | 0.272 | 0.004 |
| FIS1012a | 13 | 2.155 | 0.020 | 0.123 | 0.007 | 0.271 | 0.004 |

b. SNP (4193 loci, one random SNP/ chromosome, N=13 per population)

| Pop ID | N | No. of Alleles (mean) | No. of Alleles (SE) | No. Private Alleles (mean) | No. Private Alleles (SE) | $H_e$ (mean) | $H_e$ (SE) |
|---|---|---|---|---|---|---|---|
| ***Wild Strains*** | | | | | | | |
| *Hudson River* | | | | | | | |
| PM1012s | 13 | 1.341 | 0.007 | 0.021 | 0.002 | 0.073 | 0.002 |
| TPZ0713a | 13 | 1.337 | 0.007 | 0.020 | 0.002 | 0.075 | 0.002 |
| IRV1012s | 13 | 1.351 | 0.007 | 0.021 | 0.002 | 0.074 | 0.002 |
| HH0512a | 13 | 1.340 | 0.007 | 0.022 | 0.002 | 0.074 | 0.002 |
| HH0812s | 13 | 1.347 | 0.007 | 0.027 | 0.002 | 0.073 | 0.002 |
| HH0912s | 13 | 1.342 | 0.007 | 0.021 | 0.002 | 0.073 | 0.002 |
| HH1013a | 13 | 1.353 | 0.007 | 0.023 | 0.002 | 0.075 | 0.002 |
| HH1013s | 13 | 1.351 | 0.007 | 0.023 | 0.002 | 0.075 | 0.002 |
| PRA1013a | 13 | 1.355 | 0.007 | 0.027 | 0.002 | 0.074 | 0.002 |
| *East River* | | | | | | | |
| SV0512a | 13 | 1.349 | 0.007 | 0.018 | 0.002 | 0.075 | 0.002 |
| SV0812s | 13 | 1.344 | 0.007 | 0.019 | 0.002 | 0.075 | 0.002 |
| SV0912s | 13 | 1.344 | 0.007 | 0.018 | 0.002 | 0.075 | 0.002 |
| SV1013a | 13 | 1.343 | 0.007 | 0.020 | 0.002 | 0.074 | 0.002 |
| RI812s | 13 | 1.346 | 0.007 | 0.020 | 0.002 | 0.075 | 0.002 |
| WFM812s | 13 | 1.336 | 0.007 | 0.017 | 0.002 | 0.074 | 0.002 |
| AB812s | 13 | 1.346 | 0.007 | 0.019 | 0.002 | 0.074 | 0.002 |
| NTC1013a | 13 | 1.344 | 0.007 | 0.018 | 0.002 | 0.074 | 0.002 |
| ***Aquaculture*** | | | | | | | |
| FIS1013a | 13 | 1.224 | 0.006 | 0.002 | 0.001 | 0.063 | 0.002 |
| FIS1012a | 13 | 1.207 | 0.006 | 0.003 | 0.001 | 0.060 | 0.002 |

**Table IV.** Analysis of Molecular Variance (AMOVA) among 5 samples from Hastings on Hudson, among 4 samples from Soundview, and among 17 population samples from 10 different locations. Life history (adult vs spats), temporal replicates (2012 vs 2013), and two regional "groups" (Hudson vs East) were tested, respectively (a). The aquaculture strains were then added and compared against the wild population samples (b).

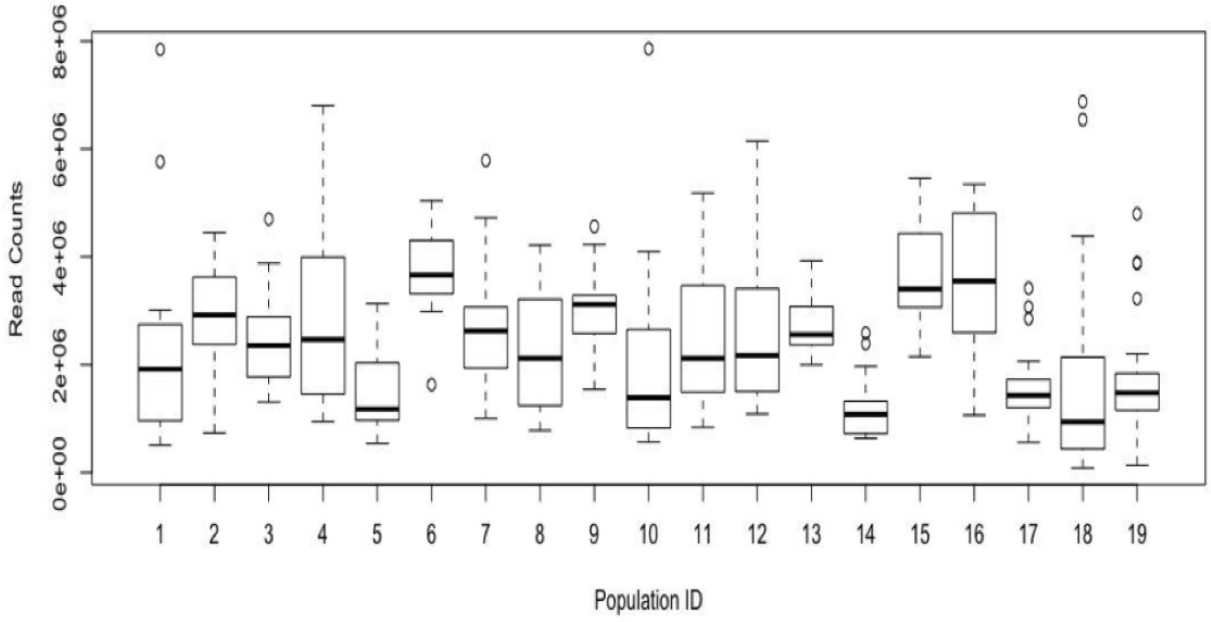
a) 5HH and 4SV populations (Life History, Temporal Replicates)

| **Source of variation** | **df** | **Proportion of variation** |
|---|---|---|
| *Hastings on the Hudson (HH)* | | |
| Among groups (adults vs spats) | 1 | 0.488 (P<0.001)* |
| Among populations within groups (2012 vs 2013) | 3 | 0.002 (P=0.41) |
| Within populations | 103 | 0.489 (P<0.001)* |
| *Soundview(SV)* | | |
| Among groups (adults vs spat) | 1 | 0.554 (P<0.001)* |
| Among populations within groups (2012 vs 2013) | 2 | 0.002 (P= 0.343) |
| Within populations | 78 | 0.555 (P<0.001)* |
| *Hudson vs East River* | | |
| Among groups (Hudson vs East River) | 1 | 0.009(P>0.05) |
| Among populations within groups (2012 vs 2013) | 15 | 0.002 (P>0.05) |
| Within populations | 332 | 0.011 (P<0.01) |

b) 17+2 populations (Wild vs Aquaculture)

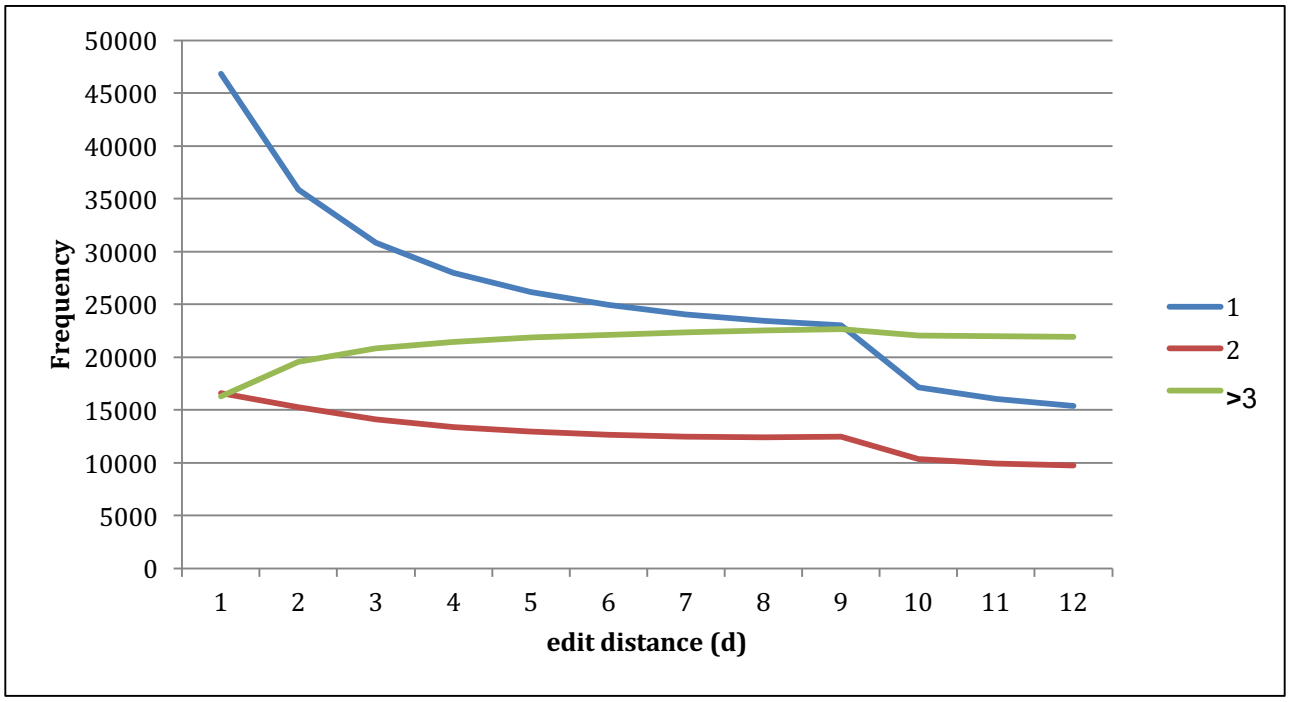
| **Source of variation** | **df** | **Proportion of variation (HH)** |
|---|---|---|
| Among groups (wild vs aquaculture) | 1 | 0.339 (P<0.001)* |
| Among populations within groups (Hudson, East, FIS2012, Fis 2013) | 3 | 0.001 (P=0.41) |
| Within populations | 92 | 0.340 (P<0.001)* |

## Supplemental Figures



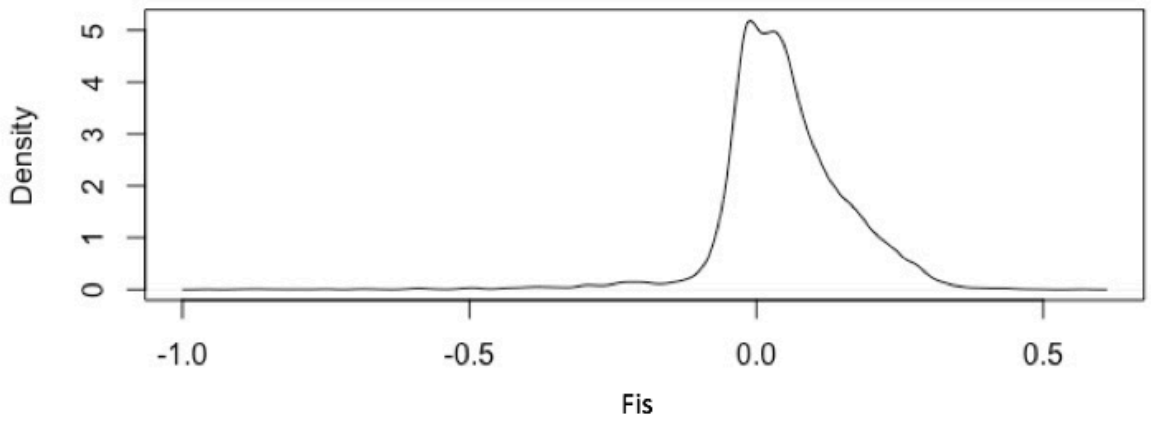

**Supplemental Figure 1: Post QC Read Counts Data.**

The distribution of read counts among loci after the QC (post *process_radtags*, but prior to the Fis whitelist filtering) in the final analyzed 17+2-sample data sets. The X axis numbers correspond to population numbers in Table I. Part of the initial QC was to remove individuals with fewer than 500,000 (0.5 million) reads.




**Supplemental Figure 2:**

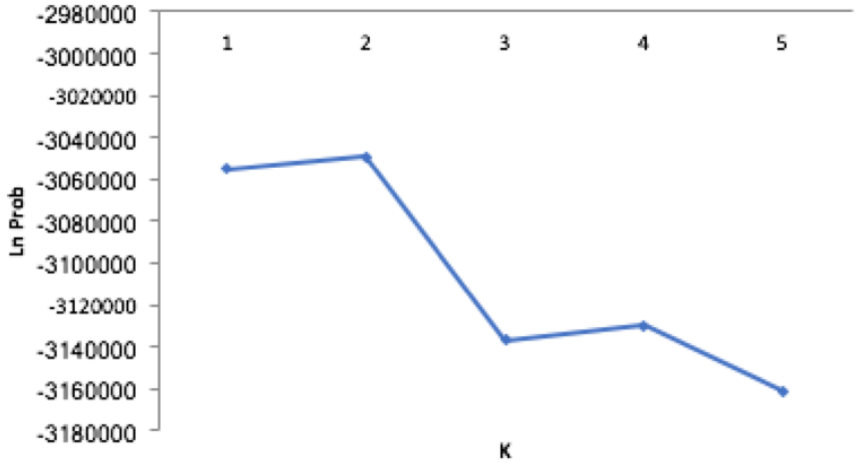
Three subsampled fasta files (1,2,>3) were used to test the edit distance. Fasta file 1 and 2 represent the homozygous files and the file >3 represent the heterozygous file. A dramatic decrease in homozygosity is shown between d9 and d10.



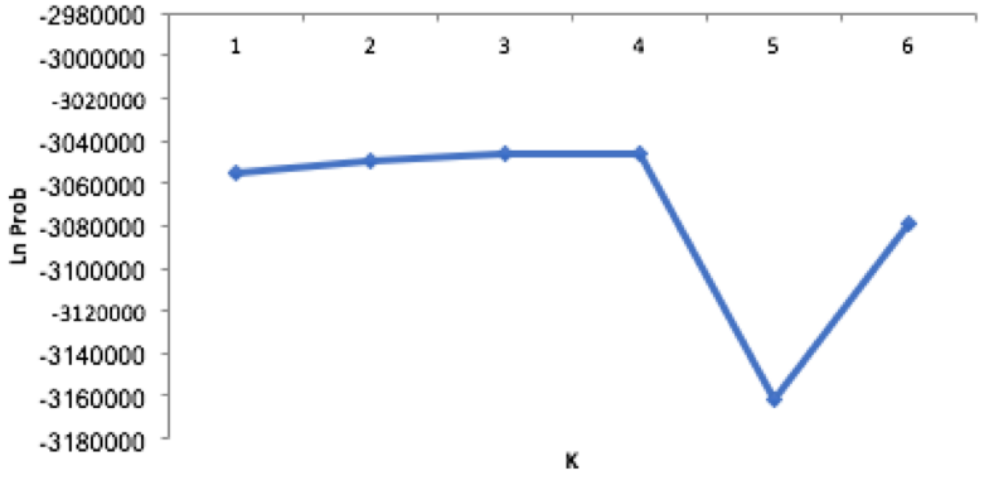

**Supplemental Figure 3: Fis Distribution for 17-wild dataset**

Treating all 349 wild individuals as a single sample, the Fis distribution among 4208 haplotype loci is shown for the 17-wild data after the custom QC filtering. The mean Fis was 0.039 with maximum Fis 0.56 and the minimum -0.956. A whitelist of loci was made based on this Fis distribution (-0.5<Fis <0.5) to generate final data from STACKS

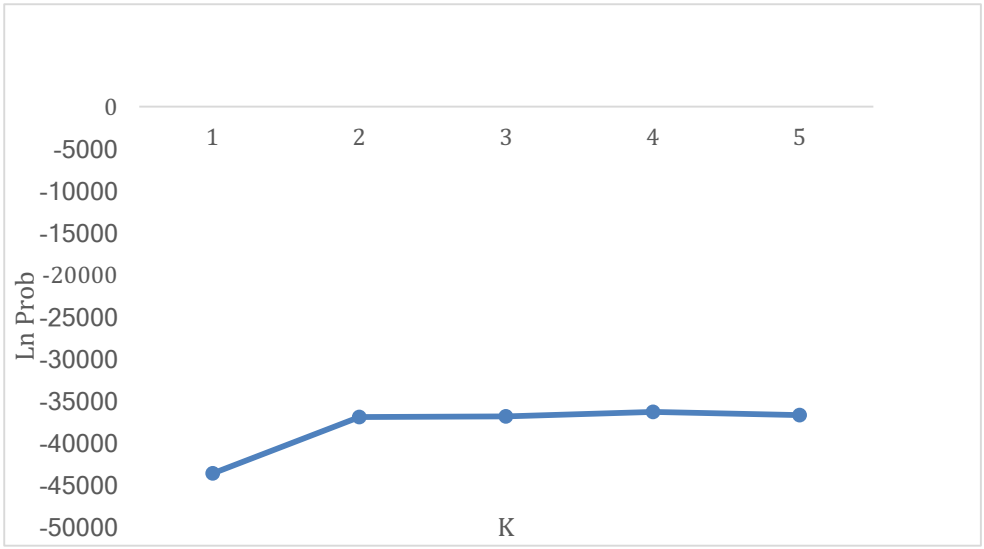
**a)**


-2980000
-3000000
-3020000
-3040000
-3060000
-3080000
-3100000
-3120000
-3140000
-3160000
-3180000
Ln Prob
1
2
3
4
5
K

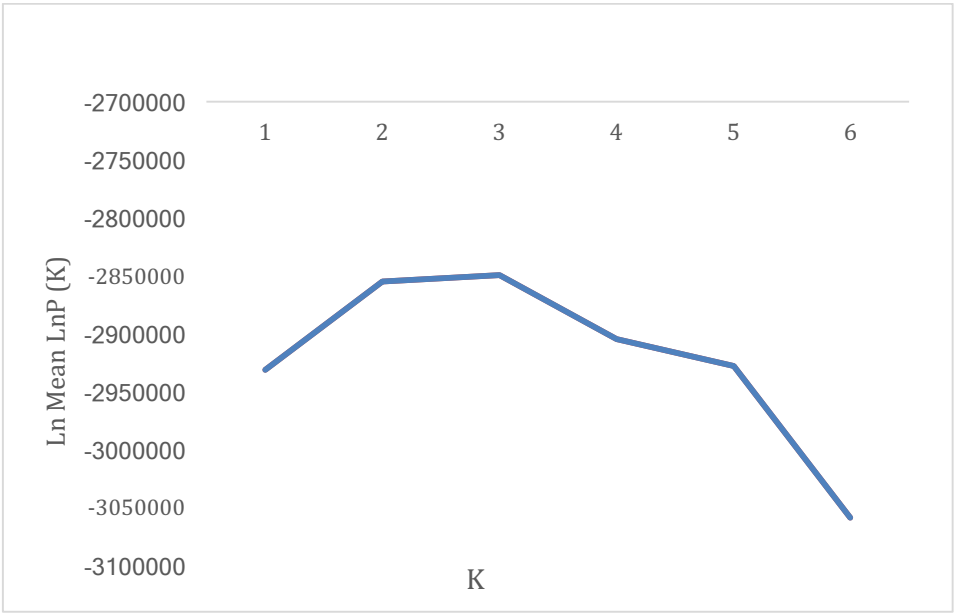
**b)**


-2980000
-3000000
-3020000
-3040000
-3060000
-3080000
-3100000
-3120000
-3140000
-3160000
-3180000
Ln Prob
1
2
3
4
5
6
K

**c )**



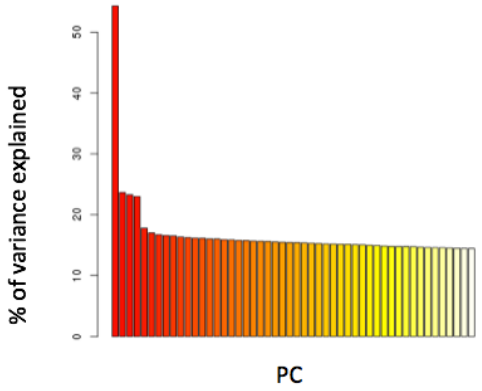

**d )**



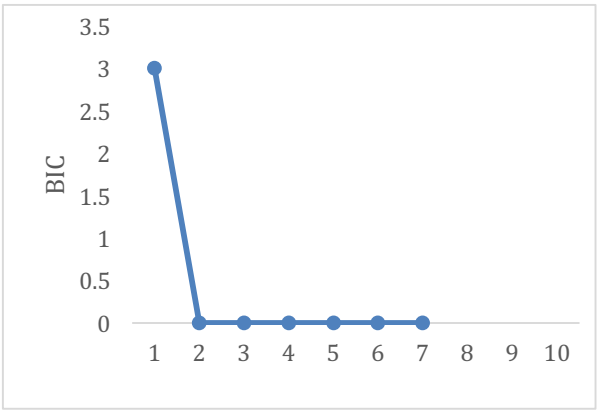

**Supplemental Figure 4: Mean Log Likelihood for Number of Populations, K**

Based on a total of three independent runs of STRUCTURE, the mean log likelihood support is shown for each K value tested with locality prior on. a) 17-wild sampleset with 4184 haplotype loci, N=349, b) 17-wild sampleset without locality prior on, c) 17-wild sampleset with 51 haplotype outlier loci, N=349, and d)19 wild/aquaculture mixed sampleset with 3655 haplotype loci, N=392

a) The PCA eigenvalues.




b) Cumulutative variance explained by number of clusters.




**Supplemental Figure 5: PCA eigenvalues and the K-means clustering results**

The PCA eigenvalues indicated that the first two principal components can explain most of the variance among the data (a). The proportion of variance between PC1-PC50 indicated that retaining ~5 PC would probably explain a substantial amount of the variation among clusters. The K-means clustering algorithm for $K = 1$ to $K = 10$ on the 17pop haplotype dataset indicated the sudden drop between K=3 and K=4. This further indicated that retaining at least 7 (conservative side) clusters for the DAPC analysis would provide helpful summary of the data by the lowest BIC (b). These results below were not informative with respect to the admixed level in individuals in the clusters, thus these are not includeed in the results section.

\

**Supplemental Table 1: Annotation for the 51 outlier loci compared against *C.gigas* reference genome (Eierman and Hare 2014).**

| Loci ID | Stack ID | Annotation from BLASTn | Accession number |
|---|---|---|---|
| 1 | 20988 | TPA_asm: Oryzias latipes strain Hd-rR, complete genome assembry chromosome 20. | HF933226.1 |
| 2 | 21660 | Crassostrea gigas ankyrin repeat domain-containing protein 50-like (LOC105319099), transcript variant X2, mRNA | XM_011416508.2 |
| 3 | 21980 | PREDICTED: Crassostrea gigas low-density lipoprotein receptor-related protein 4 (LOC105322227), mRNA | XM_011420816.2 |
| 4 | 21980 | PREDICTED: Crassostrea gigas low-density lipoprotein receptor-related protein 4 (LOC105322227), mRNA | XM_011420816.2 |
| 5 | 22049 | PREDICTED: Homo sapiens microfibrillar associated protein 3 like (MFAP3L), transcript variant X8, mRNA | XM_017008870.1 |
| 6 | 22049 | PREDICTED: Homo sapiens microfibrillar associated protein 3 like (MFAP3L), transcript variant X8, mRNA | XM_017008870.1 |
| 7 | 12481 | PREDICTED: Crassostrea gigas factor VIII intron 22 protein (LOC105345952), mRNA | XM_011454337.2 |
| 8 | 395 | Levyella sp. Marseille-P3170 strain Marseille-P3170T genome assembly, chromosome: contig00001 | LT635480.1 |
| 9 | 2765 | PREDICTED: Crassostrea gigas dual oxidase maturation factor 1 (LOC105318698), transcript variant X4, mRNA | XM_011415933.2 |
| 10 | 2765 | PREDICTED: Crassostrea gigas dual oxidase maturation factor 1 (LOC105318698), transcript variant X4, mRNA | XM_011415933 |
| 11 | 20456 | PREDICTED: Crassostrea gigas metal transporter CNNM4-like (LOC105340238), mRNA | XM_011446195.2 |
| 12 | 15699 | PREDICTED: Crassostrea gigas metal transporter CNNM4-like (LOC105340238), mRNA | XM_011446195.2 |
| 13 | 5573 | PREDICTED: Crassostrea gigas monoamine oxidase A (Maoa), transcript variant X2, mRNA | XM_011454195.2 |
| 14 | 10648 | PREDICTED: Crassostrea gigas T-box brain protein 1 (LOC105333225), transcript variant X2, mRNA | XM_011436089.2 |
| 15 | 18787 | PREDICTED: Crassostrea gigas transcription termination factor 2 (LOC105339697), mRNA | XM_011445367.2 |
| 16 | 18787 | PREDICTED: Aplysia californica putative mediator of RNA polymerase II transcription subunit 26 (LOC101855378), mRNA | XM_005109539.2 |
| 17 | 18787 | Nippostrongylus brasiliensis genome assembly N_brasiliensis_RM07_v1_5_4, scaffold NBR_scaffold0000122 | LM433505.1 |
| 18 | 8815 | Arthroderma benhamiae CBS 112371 hypothetical protein, mRNA | XM_003009964.1 |
| 19 | 7247 | PREDICTED: Crassostrea gigas cohesin subunit SA-1-like (LOC105338470), transcript variant X4, mRNA | XM_011443618.2 |
| 20 | 7247 | PREDICTED: Crassostrea gigas cohesin subunit SA-1-like (LOC105338470), transcript variant X4, mRNA | XM_011443618.2 |
| 21 | 17448 | Solanum lycopersicum chromosome ch05, complete genome | HG975517.1 |
| 22 | 2680 | PREDICTED: Crassostrea gigas E3 ubiquitin-protein ligase CBL (LOC105327949), mRNA | XM_020067105.1 |
| 23 | 21200 | PREDICTED: Aplysia californica spermatogenesis-associated protein 13-like (LOC101858536), mRNA | XM_013090038.1 |
| 24 | 21200 | PREDICTED: Aplysia californica spermatogenesis-associated protein 13-like (LOC101858536), mRNA | XM_013090038.1 |
| 25 | 7116 | PREDICTED: Lepisosteus oculatus tensin 2 (tns2), transcript variant X7, mRNA | XM_015344216.1 |
| 26 | 2680 | PREDICTED: Crassostrea gigas E3 ubiquitin-protein ligase CBL (LOC105327949), mRNA | XM_020067105.1 |
| 27 | 2912 | PREDICTED: Crassostrea gigas serine/arginine repetitive matrix protein 2 (LOC105325619), transcript variant X6, mRNA | XM_011425270.2 |
| 28 | 8047 | Crassostrea gigas clone CG_Ba50F9, complete sequence | GU207412.1 |
| 29 | 138 | Capnocytophaga haemolytica strain CCUG 32990, complete genome | CP014227.1 |